

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی کامپیوتر و فناوری اطلاعات

گروه هوش مصنوعی

تشخیص ارقام و حروف تایپی انگلیسی بدون محدودیت در اندازه و نوع فونت به کمک شبکه عصبی نگاشت خودسازمانده

دانشجو: نجمه صمدیانی

استاد راهنما:

دکتر حمید حسن پور

استاد مشاور:

مهندس سکینه اسدی امیری

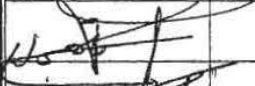
پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

تیر ۱۳۹۳

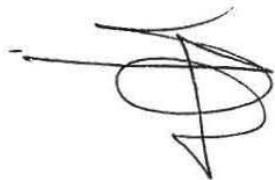
باسمه تعالی
 دانشکده: مهندسی کامپیوتر و فناوری اطلاعات
 گروه: هوش مصنوعی

پایان نامه کارشناسی ارشد: خانم نجمه صمدیانی
 تحت عنوان: تشخیص ارقام و حروف تایپی انگلیسی بدون محدودیت در اندازه و نوع فونت به کمک شبکه
 عصبی نگاشت خودسازمانده

در تاریخ ۹۳/۴/۱۷ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مورد ارزیابی و با درجه عالی مورد پذیرش قرار
 گرفت.

عضو هیات داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استاد راهنما	دکتر حمید حسن پور	استاد	
۲- استاد مشاور	مهندس اسدی امیری	دانشجوی دکتری	
۳- نماینده شورای تحصیلات تکمیلی	مهندس فرهادی	مربی	
۴- استاد ممتحن	دکتر شاه بهرامی	دانشیار	
۵- استاد ممتحن	دکتر فردوسی	استادیار	

رئیس دانشکده:



تعهد نامه

اینجانب نجمه صمدپانی دانشجوی دوره کارشناسی ارشد رشته هوش مصنوعی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهرود نویسنده پایان نامه تشخیص ارقام و حروف تایی انگلیسی بدون محدودیت در اندازه و نوع فونت به کمک شبکه عصبی نگاشت خودسازمانده تحت راهنمایی دکتر حمید حسن پور متعهد می شوم.

- تحقیقات در این پایان نامه 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 پایان نامه 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 می باشد 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 پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، 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

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 می باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

تقدیم بہ

مہربانان عزیز تر از جانم

پدر و مادر دلسوزم

شکر و قدردانی

حمد و سپاس خدای را، آن نخستین بی پیشین را و آن آخرین بی پسین را، او را
سپاس گوئیم، چنان سپاسی که چون در میان سپاسگزارانش زیستن کیریم، همواره با ما
باشد و ابواب علم ربوبیت خویش را به روی ما بکشد.

لازم می دانم از راهبانی های بی دریغ استاد فرزانه، جناب آقای دکتر حمید حسن پور،
شکر و قدردانی نمایم؛ هر چند، این سپاس، جوابگوی تمام عیار هدایت های بی شائبه ی
دور و نزدیک ایشان نیست.

هم‌چنین، مشورت‌های گرانمایه‌ی خانم مهندس سکینه اسدی امیری، استاد مشاور
این رساله را ارج می‌نم و قدردان نکته‌سنجی‌های سودمند ایشان، هستم.
از سایر استادان گرانقدر دانشکده، جناب آقای دکتر زاهدی، مدیر محترم گروه
جناب آقای دکتر ابولقاسمی و سرکار خانم دکتر مشایخی، کمال تشکر را دارم و
سپاسگزار زحمات منشی محترم گروه، جناب آقای علیرضا کریمی می‌باشم.

چکیده

بازشناسی نوری حروف در حوزه‌های گوناگون مانند تشخیص پلاک خودرو، امنیت (مانند تصدیق گذرنامه)، بارکد و سیستم‌های بانکی (چک)، نقش مهمی ایفا می‌کند و تا کنون روش‌ها و سیستم‌های متعددی برای تشخیص کاراکترها ارائه شده است. هدف این تحقیق، ارائه‌ی روشی می‌باشد که بتواند ارقام و حروف تاییبی یک زبان را بدون محدودیت در نوع فونت و اندازه، شناسایی نماید.

از مهم‌ترین چالش‌های سیستم‌های بازشناسی موجود، این است که دارای تعداد زیادی نمونه‌ی آموزشی و در نتیجه حجم بالایی از محاسبات هستند؛ زیرا هنگام آموزش هر یک از کاراکترها به چندین

نمونه‌ی آموزشی از آن کاراکتر، نیاز دارد. همچنین اگر این سیستم‌ها برای تشخیص کاراکترها با تنوع فونت‌های بیشتری توسعه یابند؛ خطای تشخیص، افزایش می‌یابد. بنابراین به سیستم بازشناسی‌ای نیاز داریم که قادر باشد علاوه بر کم بودن تعداد نمونه‌های آموزشی، در مقابل تغییر فونت، مقاوم بوده و خطای تشخیص آن در ازای افزودن فونت‌های جدید، تغییر آن‌چنانی نداشته باشد.

در این پایان‌نامه، به بازشناسی ارقام و حروف تاییپی انگلیسی می‌پردازیم. کاراکترهای تاییپی در فونت‌ها و اندازه‌های مختلف نوشته می‌شوند که سبب می‌گردد شکل نوشتاری کاراکترهای مشابه با یکدیگر متفاوت باشد؛ بنابراین به یک سیستم بازشناسی با استخراج ویژگی‌های مناسب از کاراکترها، نیاز است تا بتواند با درنظر گرفتن این تفاوت‌ها، کاراکترهای مشابه را تشخیص دهد. در این تحقیق، با استفاده از یک بردار ویژگی ساده- مجموع مقادیر پیکسل‌های تک تک سطرها و ستون‌ها- و یک شبکه‌ی عصبی نگاشت خودسازمانده بدون ساختار پیچیده، توانستیم ارقام تاییپی انگلیسی را در ۳۰ نوع فونت و ۱۱ اندازه‌ی متفاوت با دقت ۹۹,۴۷٪ و حروف تاییپی انگلیسی را در ۲۴ نوع فونت و ۱۱ اندازه‌ی مختلف با دقت ۹۷,۷۳٪، دسته‌بندی کنیم.

شبکه‌ی عصبی نگاشت خودسازمانده، معمولاً از فاصله‌ی اقلیدسی برای مقایسه‌ی نمونه‌های یک کلاس استفاده می‌کند اما در این پایان‌نامه به جای استفاده از فاصله‌ی اقلیدسی، معیار شباهت برای مقایسه‌ی بین بردار مشخصه‌ی رقم ورودی و وزن نرون‌های مختلف، به کار برده می‌شود. همچنین، به جای آموزش چندین نمونه از فونت‌ها و اندازه‌های مختلف، تنها یک یا دو نمونه برای آموزش در شبکه‌ی عصبی، مورد استفاده قرار می‌گیرد.

کلمات کلیدی: شناسایی کاراکتر، معیار شباهت، شبکه‌ی عصبی نگاشت خودسازمانده، استخراج ویژگی.

فهرست مطالب

صفحه	عنوان
ک.....	فهرست شکل ها
ن.....	فهرست جداول
۱.....	فصل اول: سیستم های بازشناسی نوری حروف
۲.....	۱-۱- مقدمه

۲-۱- معرفی سیستم‌های بازشناسی نوری حروف.....	۲
۳-۱- ضرورت تحقیق.....	۳
۴-۱- تاریخچه‌ی سیستم‌های « بازشناسی نوری حروف».....	۴
۵-۱- انواع بازشناسی نوری حروف.....	۶
۱-۵-۱- از نظر ورودی.....	۶
۱-۵-۲- از نظر نحوه‌ی ورود اطلاعات.....	۸
۶-۱- معرفی بخش‌های مختلف یک سیستم بازشناسی نوری حروف.....	۹
۱-۶-۱- پیش‌پردازش.....	۹
۲-۶-۱- قطعه‌بندی.....	۱۲
۳-۶-۱- استخراج ویژگی‌ها.....	۱۳
۴-۶-۱- طبقه‌بندی و شناسایی.....	۱۴
۷-۱- ارزیابی سیستم‌های بازشناسی نوری حروف.....	۱۴
۸-۱- خلاصه و نتیجه‌گیری.....	۱۵
فصل دوم: مروری بر کارهای پیشین.....	۱۶
۱-۲- مقدمه.....	۱۷
۲-۲- مروری بر روش‌های موجود بازشناسی نوری حروف.....	۱۷
۳-۲- بازشناسی نوری حروف بدون محدودیت در اندازه و نوع فونت.....	۲۳
۴-۲- نتیجه‌گیری.....	۲۵
فصل سوم: پیاده‌سازی سیستم پیشنهادی.....	۲۷
۱-۳- مقدمه.....	۲۸
۲-۳- سیستم شناسایی پیشنهادی.....	۲۸

۲۸.....	۳-۲-۱- پایگاه داده
۲۹.....	۳-۲-۲- پیش پردازش
۳۰.....	۳-۲-۳- استخراج مشخصه ها
۴۵.....	۳-۲-۴- طبقه بندی و شناسایی
۴۵.....	۳-۲-۴-۱- معیار شباهت
۴۹.....	۴-۲-۲- شبکه عصبی نگاشت خودسازمانده
۵۳.....	۳-۲-۵- انتخاب ویژگی ها با الگوریتم وراثتی
۵۵.....	۳-۳- نتیجه گیری
۵۷.....	فصل چهارم: ارزیابی عملکرد و تحلیل نتایج
۵۸.....	۴-۱- مقدمه
۵۹.....	۴-۲- نتایج
۵۹.....	۴-۲-۱- نتایج حاصل از دسته بندی ارقام انگلیسی
۶۲.....	۴-۲-۲- نتایج حاصل از دسته بندی حروف انگلیسی
۷۰.....	۴-۲-۲- نتایج حاصل از دسته بندی ارقام فارسی
۷۲.....	۴-۳- نتیجه گیری
۷۴.....	فصل پنجم: نتیجه گیری و پیشنهادات
۷۵.....	۵-۱- نتیجه گیری
۷۶.....	۵-۲- پیشنهادات
۷۹.....	پیوست ۱
۸۴.....	پیوست ۲
۹۱.....	فهرست مراجع

فهرست شکل‌ها

عنوان	صفحه
شکل ۱-۱: فرآیند تبدیل یک متن اسکن شده به یک فایل متنی قابل ویرایش با استفاده از بازشناسی نوری حروف.	۳
شکل ۲-۱: تقسیم‌بندی سیستم‌های بازشناسی نوری حروف.	۶
شکل ۳-۱: الف) نمونه‌ای از یک متن دست‌نویس گسسته، ب) نمونه‌ای از یک متن دست‌نویس پیوسته.	۷
شکل ۴-۱: نمودار بلوکی دیاگرام یک سیستم بازشناسی نوری حروف.	۹

- شکل ۱-۵: الف) تصویری از یک متن اسکن شده به صورت کج، ب) پس از اصلاح کجی آن..... ۱۱
- شکل ۱-۶: قطعه‌بندی یک متن دست‌نوشته با استفاده از روش غیر صریح..... ۱۲
- شکل ۱-۷: قطعه‌بندی تصویر ورودی با استفاده از روش صریح..... ۱۳
- شکل ۱-۸: مثال‌هایی از تصاویر اسناد با پس‌زمینه‌های طرح‌دار و غیریکنواخت..... ۱۵
- شکل ۳-۱: نتیجه‌ی اسکن شده‌ی یک نمونه‌ی تایپی عدد هفت، چرخش ۴۵ درجه‌ی آن خلاف عقربه‌های ساعت و تشکیل مستطیل محاطی آن به منظور تشکیل بردار مشخصه..... ۳۰
- شکل ۳-۲: نمونه‌ای از رقم هفت با فونت Times New Roman و بردار مشخصه‌ی آن..... ۳۲
- شکل ۳-۳: شباهت نیمه‌ی بالایی دو رقم 0 و 9، الف) قبل و ب) بعد از چرخش ۴۵ درجه در جهت خلاف عقربه‌های ساعت..... ۳۳
- شکل ۳-۴: الف) سری زمانی حاصل از بردار ویژگی ارقام ۰ و ۹ قبل از چرخش، ب) بعد از چرخش..... ۳۳
- شکل ۳-۵: سری زمانی حاصل از بردار ویژگی استخراج شده برای ارقام الف) ۰، ب) ۱، ج) ۲، چ) ۳، ح) ۴، خ) ۵، د) ۶، ذ) ۷، ر) ۸، ز) ۹، قبل از چرخش..... ۳۴
- شکل ۳-۶: سری زمانی حاصل از بردار ویژگی استخراج شده برای ارقام الف) ۰، ب) ۱، ج) ۲، چ) ۳، ح) ۴، خ) ۵، د) ۶، ذ) ۷، ر) ۸، ز) ۹ پس از چرخش..... ۳۵
- شکل ۳-۷: نمونه‌ای از کاراکتر A با فونت Arial و ویژگی‌های نیمه‌ی بالایی تصویر..... ۳۷
- شکل ۳-۸: نمونه‌ای از کاراکتر A با فونت Arial و ویژگی‌های نیمه‌ی پایینی تصویر..... ۳۸
- شکل ۳-۹: الف) سری زمانی ۱۶۵ تایی حاصل از بردار ویژگی نمونه کاراکتر A نشان داده شده در شکل‌های ۳-۷ و ۳-۸، ب) بردار ویژگی نمونه کاراکتر A مذکور..... ۳۹
- شکل ۳-۱۰: سری زمانی حاصل از بردار ویژگی استخراج شده برای حروف A تا F را قبل از چرخش ۴۵ درجه..... ۳۹
- شکل ۳-۱۱: سری زمانی حاصل از بردار ویژگی استخراج شده برای حروف A تا Z پس از چرخش..... ۴۲
- شکل ۳-۱۲: سری زمانی حاصل از بردار ویژگی استخراج شده برای حروف a تا z پس از چرخش..... ۴۴

- شکل ۳-۱۳: سری زمانی حاصل از دو نمونه از رقم ۲ با فونت Times new roman ۴۵
- شکل ۳-۱۴: بردار ویژگی نمونه‌ای از حرف A در اندازه ۲۴. الف) فونت Arial، ب) فونت Cambria ۴۷
- شکل ۳-۱۵: نمودار اصلی سری زمانی رقم ۳، نمودار با علامت * وزن برنده‌ی کلاس رقم ۳ و نمودار با علامت + وزن برنده رقم ۲ ۴۷
- شکل ۳-۱۶: نمودار اصلی سری زمانی حرف A، نمودار با علامت * وزن برنده‌ی کلاس حرف B و نمودار با علامت + وزن برنده حرف A ۴۸
- شکل ۳-۱۷: یک شبکه SOM ۵۱
- شکل ۳-۱۸: نمونه‌های مختلف نوشتاری برخی از حروف ۵۳
- شکل ۴-۱: نمودارهای بهترین برازندگی و متوسط برازندگی در ۲۰ تکرار الگوریتم برای کاهش ویژگی‌های سیستم شناسایی ارقام ۶۱
- شکل ۴-۲: نمونه‌هایی از حروف a, g, l و y که به دو صورت کاملاً متفاوت در فونت‌های مختلف نوشته می‌شوند ۶۳
- شکل ۴-۳: نمودارهای بهترین برازندگی و متوسط برازندگی در ۵ تکرار الگوریتم برای کاهش ویژگی حروف بزرگ از ۱۶۵ به ۸۰ ویژگی ۶۶
- شکل ۴-۴: نمودارهای بهترین برازندگی و متوسط برازندگی در ۲۰ تکرار الگوریتم برای کاهش ویژگی‌های حروف بزرگ از ۸۰ به ۳۰ ویژگی ۶۷
- شکل ۴-۵: نمودارهای بهترین برازندگی و متوسط برازندگی در ۲۰ تکرار الگوریتم برای کاهش ویژگی‌های حروف کوچک ۶۷
- شکل ۴-۶: نمونه‌ای از حروف A تا F به صورت نازک شده و بردارهای ویژگی قبل و بعد از نازک‌سازی آنها ۷۰
- شکل ۴-۷: انواع شکل‌های ۴ و ۵ و ۶ در زبان فارسی ۷۱

فهرست جداول

عنوان	صفحه
جدول ۱-۳: استفاده از معیار شباهت Jensen برای ارزیابی شباهت بین بردارهای مشخصه‌ی استخراج شده از ارقام مختلف.	۵۰
جدول ۲-۳: استفاده از معیار شباهت Jensen برای ارزیابی شباهت بین بردارهای مشخصه‌ی استخراج شده از حروف مختلف.	۵۶
جدول ۱-۴: حالات مختلف یک پیش‌بینی دو کلاسه.	۵۸
جدول ۲-۴: دقت روش پیشنهادی بر روی داده‌های ارقام.	۶۰

- جدول ۳-۴: نتیجه‌ی دسته‌بندی ارقام با روش پیشنهادی. ۶۰.....
- جدول ۴-۴: مقایسه‌ی روش پیشنهادی شناسایی ارقام با سایر روش‌های موجود در این حوزه. ۶۲.....
- جدول ۵-۴: دقت روش پیشنهادی بر روی حروف بزرگ. ۶۴.....
- جدول ۶-۴: نتیجه دسته‌بندی روش پیشنهادی بر روی حروف بزرگ. ۶۴.....
- جدول ۷-۴: دقت روش پیشنهادی بر روی حروف کوچک. ۶۵.....
- جدول ۷-۴: نتیجه دسته‌بندی روش پیشنهادی بر روی حروف کوچک. ۶۵.....
- جدول ۸-۴: مقایسه‌ی روش پیشنهادی تشخیص حروف با سایر روش‌های موجود در این حوزه. ۶۸.....
- جدول ۹-۴: دقت روش پیشنهادی بر روی ارقام فارسی. ۷۱.....
- جدول ۱۰-۴: نتیجه دسته‌بندی روش پیشنهادی برای شناسایی ارقام فارسی. ۷۲.....
- جدول ۱۱-۴: مقایسه‌ی روش پیشنهادی با سایر روش‌های موجود شناسایی ارقام فارسی. ۷۲.....

فصل اول

سیستم‌های بازشناسی نوری حروف

۱-۱- مقدمه

جهان، در حال حرکت به سوی استفاده‌ی گسترده از کامپیوترها و ابزارهای دیجیتالی برای دسترسی به اطلاعات است. تعداد بی‌شماری از افراد، از اینترنت به عنوان یک منبع مطمئن برای دستیابی به اطلاعات، جست‌وجو برای مقالات، خواندن اخبار و به‌روزرسانی اطلاعات خود استفاده می‌کنند. بنابراین، باید منابع اطلاعاتی مانند مجلات، کتاب‌ها و سایر مطبوعات، دیجیتالی شوند و به صورت الکترونیکی در اختیار افراد قرار گیرند؛ پس نیاز به استفاده از روش‌های سریع و تا حد زیادی مطمئن موسوم به «ا.سی. آر» یا «بازشناسی نوری حروف» به منظور وارد نمودن اطلاعات موجود در اسناد، مدارک، کتاب‌ها و سایر مکتوبات تایپی و حتی دست‌نوشته به داخل رایانه هستیم.

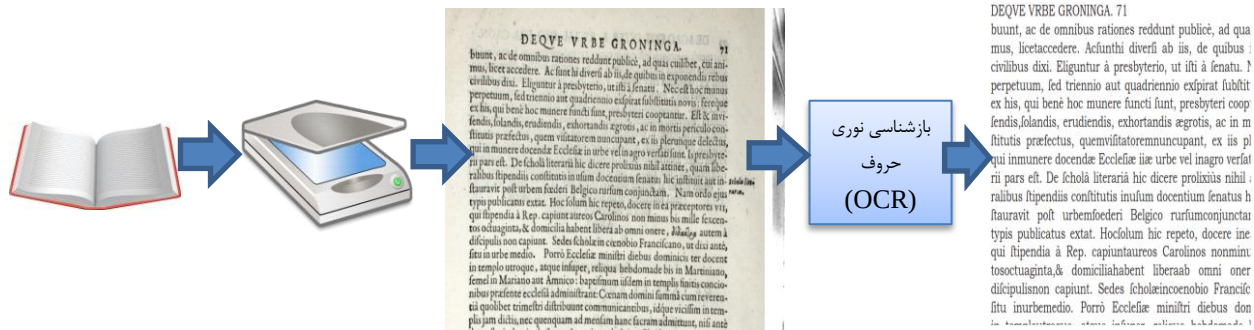
در این فصل، ابتدا به شرح کامل سیستم‌های بازشناسی نوری حروف و تاریخچه‌ی آنها می‌پردازیم. سپس، انواع این سیستم‌ها را بیان کرده و بخش‌های مختلف یک سیستم بازشناسی نوری حروف را معرفی می‌کنیم. ضرورت تحقیق و نتیجه‌گیری، پایان‌بخش این فصل خواهند بود.

۱-۲- معرفی سیستم‌های بازشناسی نوری حروف

اصطلاح «ا.سی. آر» یا «بازشناسی نوری حروف» به تکنیک‌هایی اطلاق می‌شود که در تصاویر اسکن یا فکس شده، نواحی متنی را تشخیص می‌دهند و سپس این نواحی را به متن قابل ویرایش تبدیل می‌نمایند [۱]. این تکنولوژی، به صورت خودکار کاراکترها را با یک مکانیسم نوری بازشناسی می‌کند؛ همانطور که چشم‌های انسان، مکانیسم نوری او بوده و تصویر دیده شده توسط چشم، ورودی برای مغز انسان است. «ا.سی. آر» تکنیکی است که مانند توانایی انسان برای خواندن عمل می‌کند [۲].

تصویر یک صفحه کاغذ به صورت یک فایل با اسکنرگرافیکی، به رایانه ارسال و در آن ذخیره می‌شود. بدین ترتیب، کاربر می‌تواند با یک نرم‌افزار مناسب نمایش‌دهنده‌ی تصاویر، تصویر صفحه‌ی اسکن‌شده را بر روی نمایشگر رایانه‌ی خود ملاحظه نماید؛ اما قادر نخواهد بود که متن موجود در تصویر سند را ویرایش کند یا آن را مورد جست‌وجو قرار دهد. یک نرم‌افزار «بازشناسی نوری حروف»، تصویر اسکن‌شده را می‌خواند، محتویات آن (شامل متن، خطوط، تصاویر، جداول، ...) را شناسایی می‌نماید و

سپس با استفاده از تکنیک‌های یادگیری ماشین و بینایی کامپیوتر، آن را به یک قالب قابل ویرایش در واژه‌پردازها مانند ورد^۱ تبدیل می‌کند. شکل ۱-۱، این فرآیند را به نمایش می‌گذارد.



شکل ۱-۱: فرآیند تبدیل یک متن اسکن شده به یک فایل متنی قابل ویرایش با استفاده از بازشناسی نوری حروف.

۱-۳- ضرورت تحقیق

علی‌رغم پژوهش‌های گسترده‌ای انجام گرفته در حوزه‌ی بازشناسی نوری کاراکتر، روش‌های موجود دارای یک مشکل اساسی هستند و آن اینکه این روش‌ها به فونت، اندازه و یا حالت کاراکتر (نرمال^۲، پررنگ^۳ و ایتالیک) بسیار وابسته هستند. به خصوص اگر بخواهیم از روش‌های موجود برای شناسایی ارقامی با فونت متفاوت نسبت به نمونه‌های آموزشی استفاده نماییم، خطای تشخیص افزایش می‌یابد. بر این اساس، اندازه‌ی پایگاه داده‌ی روش‌های موجود بسیار بزرگ می‌باشد تا به ازای فونت‌های مختلف و یا اندازه‌های گوناگون نماینده‌ای در پایگاه داده هنگام آموزش داشته باشند. همچنین، با گذشت زمان بر اساس سلاقی مختلف، به تعداد فونت‌های موجود افزوده می‌شود و تنوع فونت‌ها بالا می‌رود. بنابراین روش‌های موجود «بازشناسی نوری حروف» قادر به تشخیص این فونت‌های جدید نبوده و نیازمند تغییر می‌باشند؛ اما با وجود یک سیستم بازشناسی که بتواند در مقابل تنوع فونت مقاوم باشد، با ایجاد فونت‌های جدید خللی در تشخیص نمونه‌ها به وجود نخواهد آمد. این پایان‌نامه، چنین سیستمی را ارائه می‌دهد.

^۱ Word

^۲ Regular

^۳ Bold

با آموزش تنها یک نمونه از هر یک از کاراکترها و استفاده از شباهت بین نمونه‌ها، می‌توان کاراکترهای یک دسته را تشخیص داد؛ همانطور که اگر از شخصی خواسته شود کاراکترهایی با فونت‌های جدید را شناسایی کند، از مفهوم شباهت استفاده می‌کند و نیازی ندارد که قبلاً کاراکترها را در چندین فونت دیده باشد. در واقع، او با سنجش میزان شباهت بین کاراکترها و نمونه‌هایی که از قبل دیده، می‌تواند کاراکترها با فونت‌های جدید را نیز به درستی شناسایی کند.

بنابراین، علل ضرورت این تحقیق را می‌توان در موارد زیر خلاصه کرد:

۱. مقاوم بودن نسبت به تغییر فونت.
۲. مقاوم بودن نسبت به تغییر اندازه.
۳. مقاوم بودن نسبت به حالات (نرمال، بولد، ایتالیک) کاراکتر.
۴. نیاز به تعداد کم نمونه‌های آموزشی.
۵. سرعت بیشتر بازشناسی و کم کردن بار محاسباتی.

۱-۴- تاریخچه‌ی سیستم‌های «بازشناسی نوری حروف»

از جنبه‌ی تاریخی، سیستم‌های «بازشناسی نوری حروف» تا کنون سه مرحله‌ی تکاملی را پشت سر گذاشته‌اند [۳]:

الف. مرحله‌ی تکوین (از ۱۹۰۰ تا ۱۹۸۰):

اولین اقدامات صورت گرفته در زمینه‌ی بازشناسی نوری حروف، به سال‌های اول دهه‌ی ۱۹۰۰ میلادی باز می‌گردد؛ زمانی که تیورین^۱ دانشمند روسی، توانست وسیله‌ای برای خواندن افراد مبتلا به نارسایی‌های بینایی، ثبت نماید [۴]. در سال‌های ۱۹۲۹ و ۱۹۳۳ میلادی، سیستمی جهت بازشناسی نوری حروف چاپی، ساخته شد [۵]. روش تطابق الگوی استفاده شده در این سیستم به دلیل فناوری اپتومکانیکی آن، کاربردی نبود. عملکرد این سیستم، بدین صورت بود که به هر حرف نور تابیده می‌شد و نور بازتابیده شده از حروف، از ماسک‌های مکانیکی مختلفی عبور داده می‌شدند. هرگاه نوری از ماسک عبور نمی‌کرد، حرف ورودی بازشناسی و تشخیص داده می‌شد.

¹ Tyurin

در این دوره، تحقیقات موفق‌تری بیشتر بر روی حروف و اعداد لاتین انجام گرفت که شرایط و پیش‌فرض‌های خاصی برای کاراکترهای ورودی در نظر می‌گرفتند. با این حال، مطالعات چندی نیز بر روی حروف ژاپنی، چینی، عبری، هندی، سیریلیکی، یونانی و عربی در هر دو زمینه‌ی حروف چاپی و دست‌نوشته آغاز گردید. با ظهور صفحات رقومی‌کننده در دهه‌ی ۱۹۵۰ که قادر به تشخیص مختصات حرکتی نوک یک قلم مخصوص بودند، سیستم‌های «بازشناسی نوری حروف» تجاری نیز، امکان عرضه یافتند. این نوآوری سبب شد که محققان بتوانند در زمینه‌ی بازشناسی برخط حروف دست‌نوشته، فعالیت خود را آغاز نمایند [۶].

ب. مرحله‌ی توسعه (از ۱۹۸۰ تا ۱۹۹۰):

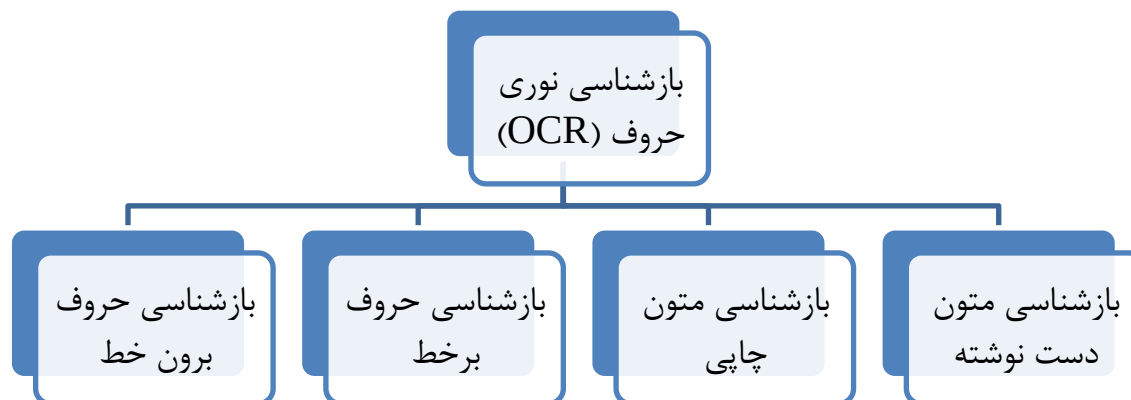
مطالعات صورت گرفته تا قبل از سال ۱۹۸۰ به دلیل فقدان سخت‌افزارهای رایانه‌ای قدرتمند و دستگاه‌های اخذ داده‌ها با مشکل همراه بودند. در این دهه به واسطه‌ی رشد فناوری اطلاعات، وضعیت بسیار مناسبی برای تحقیقات مختلف از جمله بازشناسی حروف فراهم گردید. روش‌های ساختاری به همراه روش‌های آماری در بسیاری از سیستم‌ها استفاده شدند. مروری بر تحقیقات و پیشرفت‌های حاصل‌شده در مورد «بازشناسی نوری حروف» در این برهه‌ی زمانی را می‌توان در [۵] و [۶] به ترتیب برای بازشناسی برخط و برون‌خط جستجو نمود. همچنین، روش‌های طبقه‌بندی جدیدی مانند شبکه‌های عصبی و روش‌های فازی به کار گرفته شدند.

ج. مرحله‌ی بهبود (از ۱۹۹۰ به بعد):

در این مقطع زمانی بود که با تکوین ابزارها و تکنیک‌های پردازشی جدید، پیشرفت واقعی در سیستم‌های «بازشناسی نوری حروف» محقق گردید. در اوایل دهه‌ی ۹۰، روش‌های پردازش تصویر و بازشناسی الگو با تکنیک‌های کارآمد هوش مصنوعی ادغام گشتند. محققان، الگوریتم‌های پیچیده‌ای را در بازشناسی حروف ابداع نمودند که قادر بودند داده‌های ورودی با تفکیک‌پذیری بالا را دریافت کنند و در مرحله‌ی پیاده‌سازی، محاسبات بسیار زیادی را بر روی داده‌ها انجام دهند. امروزه علاوه بر وجود رایانه‌های قدرتمندتر و تجهیزات الکترونیکی دقیق‌تر مانند اسکنرها، دوربین‌ها و صفحات رقومی‌کننده، استفاده از تکنیک‌های پردازشی مدرن و توانمند همچون شبکه‌های عصبی، مدل‌های مخفی مارکوف، منطق‌های مجموعه‌ی فازی و مدل‌های پردازش زبان طبیعی امکان‌پذیر گشته است.

۱-۵- انواع بازشناسی نوری حروف

سیستم‌های بازشناسی عمدتاً از دو منظر نوع ورودی و نحوه‌ی ورود اطلاعات، مورد بررسی و طبقه‌بندی قرار می‌گیرند.



شکل ۱-۲: تقسیم‌بندی سیستم‌های بازشناسی نوری حروف.

۱-۵-۱- از نظر ورودی

در یک تقسیم‌بندی کلی، می‌توان سیستم‌های بازشناسی نوری حروف را به لحاظ نوع الگوی ورودی به دو گروه اصلی تقسیم کرد:

الف (سیستم های بازشناسی متون چاپی

در سیستم‌های بازشناسی متون چاپی، ورودی به صورت تصویری است که از دوربین دیجیتال یا اسکنر به دست می‌آید. در فرآیند شناسایی، تصویر سند به خطوط، کلمات و حروف، شکسته می‌شوند. محدود بودن نوع فونت و جدابودن حروف در متون چاپی، الگوریتم شناسایی را ساده‌تر و عملکرد سیستم را افزایش می‌دهند. سیستم‌های بازشناسی متون چاپی، معمولاً دقت بالاتر دارند؛ زیرا کاراکترها در فونت‌های استاندارد هستند و تطابق تصاویر با نمونه‌هایی که در پایگاه داده‌های موجود می‌باشند، نسبتاً آسان است [۷].

ب (سیستم های بازشناسی متون دستنویس

متن‌های دست‌نویس به دو صورت گسسته و پیوسته وجود دارند؛ سیستم شناسایی متون دست‌نویس گسسته، شامل متونی است که در آن حروف جدا از هم، نوشته می‌شوند مانند حروفی که در فرم‌های آزمون‌ها در کادرهایی به صورت مجزا نوشته می‌شوند ولی متون دست‌نویس پیوسته در قالب نامه یا متن یا گزارش توسط شخص روی کاغذ، نوشته می‌شود. شکل ۱-۳، تصاویری از این دو نوع متن را نشان می‌دهد. متون دست‌نویس پیوسته به علت مشکلاتی از قبیل کشیده یا شکسته نوشتن بعضی از حروف یا به طور کلی نوشتاری که مطابق قواعد نگارش رسم‌الخط نیست؛ هنوز به صورت یک سیستم نرم‌افزاری جامع، درنیامده است و هم‌چنان تلاش در ایجاد سیستم شناسایی متون دست‌نویس پیوسته با دقت قابل قبول، وجود دارد. اما برای شناسایی متون دست‌نویس گسسته به علت گسسته بودن حروف و هم‌چنین متون تایی به علت آنکه عموماً از قواعد خاصی پیروی می‌کنند؛ سیستم‌های نرم‌افزاری زیادی وجود دارند که بسیاری از آنها با دقت بالایی، عملیات شناسایی را انجام می‌دهند.

- QUESTION(S) PRESENTED
- I. Whether the Arkansas Department of Corrections' no beard grooming policy violates the Religious Land Use and Institutionalized Persons Act (RLUIPA).
 - II. Whether a 1/2 inch beard would satisfy the security goals sought by the policy.
 - III. Whether the no beard grooming policy violates Petitioner's First Amendment right to practice Islam as he believes it is supposed to be practiced by the wearing of the beard.
 - IV. That the United States Court of Appeals for the Eighth Circuit has decided that the no beard grooming policy does not violate the RLUIPA, but this Court should decide the matter since it has not done so and should rule whether grooming

PERSONAL INFORMATION									
1. Claimant First Name	2. Middle Initial	3. Claimant Last Name							
JANE	J	DOE							
4. Mailing Address							5. Phone Number		
9994 ANY LANE							4105551234		
6. City	7. County	8. State	9. Zip Code						
BALTIMORE	BA	MD	21200						
10. Social Security Number	11. Sex	12. Date of Birth	13. Marital Status	14. Gross Wages Per Week	15. Paid full wages for day?				
999999999	M	02/11/97	S	100000	YES				
16. What is Your Regular Work?				17. What Was Your Work When Injured?					
SALES									

(الف)

(ب)

شکل ۱-۳: الف) نمونه‌ای از یک متن دست‌نویس گسسته، ب) نمونه‌ای از یک متن دست‌نویس پیوسته.

۱-۵-۲- از نظر نحوه‌ی ورود اطلاعات

از جنبه‌ی نحوه‌ی ورود اطلاعات، سیستم‌های «بازشناسی نوری حروف» به دو دسته‌ی سیستم‌های برخط^۱ و برون‌خط^۲ تقسیم می‌شوند. در سیستم‌های برون‌خط، دستگاه‌های ورودی، دستگاه‌هایی مانند اسکنرها و دوربین‌ها هستند درحالی که در سیستم‌های برخط، دستگاه‌های پویایی مانند تبلت‌های گرافیکی و تبلت‌های رقمی‌کننده، استفاده می‌شوند [۹].

الف) سیستم‌های برخط

ورودی سیستم بازشناسی برخط، به صورت پیوسته همزمان با نوشتن نویسنده بر روی یک صفحه رقمی‌کننده به سیستم، وارد می‌شود. بازشناسی برخط، فقط در بازشناسی دست‌نوشته، کاربرد دارد. بازشناسی برخط نوشتار به دلیل راحت‌تر بودن نوشتن از تایپ کردن، عدم امکان تایپ در بعضی مکان‌ها، عدم وجود یک صفحه کلید کامل روی رایانه‌های کوچک و سخت بودن تایپ نویسه‌ها در بعضی زبان‌ها به دلیل تعداد زیاد آنها، مورد توجه خاصی قرار گرفته است. در روش برخط از داده‌های یک بعدی، استفاده می‌شود. در این سیستم‌ها می‌توان قسمت‌های مختلف متن را ذخیره کرده، سپس عملیات شناسایی را توسط روش‌های برون‌خطی انجام داد. برای بازشناسی دست‌نوشته‌ی برخط از ویژگی‌هایی مانند مختصات مکانی حرکت قلم روی صفحه، سرعت نوشتن و میزان فشار قلم بر صفحه، استفاده می‌شود [۸].

ب) سیستم‌های برون‌خط

در سیستم بازشناسی برون‌خط، تصویر اسکن شده به عنوان ورودی در نظر گرفته می‌شود. بازشناسی برون‌خط هم شامل بازشناسی نوشتار چاپی و هم نوشتار دست‌نویس می‌شود. روش‌های شناسایی برون‌خط شامل کاربرد شبکه عصبی، منطق فازی و غیره می‌باشد که بر روی تصاویر دو بعدی، اعمال می‌شوند. در این روش به هیچ نوع وسیله‌ی نگارش خاصی نیاز نیست و تفسیر داده، مستقل از فرآیند تولید آنها، تنها بر اساس تصویر متن، صورت می‌گیرد. این روش به نحوه‌ی بازشناسی توسط انسان، شباهت بیشتری دارد.

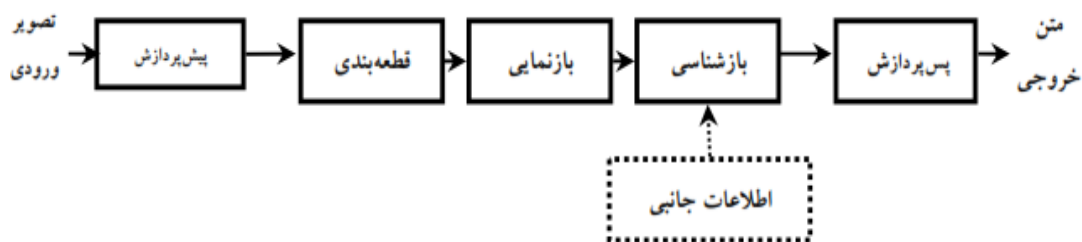
¹ Online

² Offline

سیستم‌های برون خط دقت کمتری از سیستم‌های برخط دارند. نرخ بازشناسی برای سیستم‌های برخط بالاتر از این نرخ در سیستم برون خط گزارش شده است [۹، ۱۰].

۱-۶- معرفی بخش‌های مختلف یک سیستم بازشناسی نوری حروف

شکل ۱-۴، نمودار بلوکی یک سیستم «بازشناسی نوری حروف» را نمایش می‌دهد. لازم به ذکر است که بسته به الگوریتم کلی به کار رفته و سطح انتظارات از عملکرد نرم‌فزار، ممکن است برخی سیستم‌ها فاقد یک یا چند مرحله از مراحل مذکور باشند.



شکل ۱-۴: نمودار بلوکی دیاگرام یک سیستم بازشناسی نوری حروف.

در ادامه هر یک از این بلوک‌ها مورد بررسی قرار می‌گیرند.

۱-۶-۱- پیش‌پردازش

در این مرحله، داده‌ی خام با اعمال فرآیندهایی بر روی آن، برای پردازش‌های بعدی آماده می‌شود. بنابراین، پیش‌پردازش یک مرحله‌ی اصلی است که تغییراتی را که باعث تقلیل نرخ بازشناسی می‌شود، کاهش می‌دهد. در مرحله‌ی پیش‌پردازش، اطلاعات نامرتب مانند پشت‌زمینه‌های غیر سفید تصویر و نویز، حذف می‌گردند و یک تصویر باینری با مشخصه‌های متنی باقی می‌ماند. بدین منظور مجموعه‌ای از عملگرهای پیش‌پردازش به تصویر اعمال می‌شوند که انتخاب آنها متناسب با ساختار تصویر صورت می‌گیرد [۱۱]. باینری کردن تصویر، تکنیک‌های بهبود تصویر برای حذف نویز یا بهبود کنتراست، آستانه‌گیری برای حذف پشت‌زمینه‌های پیچیده یا سایه‌دار، قطعه‌بندی صفحه برای جداسازی متن از تصویر و عملیات مورفولوژی برای بازسازی یا حذف قسمت‌های از دست رفته یا پیکسل‌های ایجاد شده در قطعه‌بندی، از جمله مواردی است که در پیش‌پردازش اعمال می‌شوند [۱۲].

باینری کردن تصویر متن: تکنیک‌های پیش‌پردازش بر روی یکی از سه صورت تصاویر رنگی، تصاویر سطح خاکستری و تصاویر باینری اعمال می‌شوند. اما به منظور کاهش حجم ذخیره‌سازی مورد نیاز و افزایش سرعت و سهولت پردازش، در اکثر کاربردها، مطلوب است که با انتخاب یک سطح آستانه، تصاویر سطح خاکستری یا رنگی را به تصاویر باینری تبدیل نمود.

تکنیک‌های بهبود تصویر: این تکنیک‌ها، کیفیت تصاویر را با حذف نویز، کاهش تاری، افزایش کنتراست و فراهم کردن جزئیات بیشتر، افزایش می‌دهند. عملگرهای فیلترینگ مکانی^۱ که برای بهبود تصاویر استفاده می‌شوند، به دو دسته‌ی پردازش نقطه‌ای^۲ و پردازش ماسک^۳ تقسیم می‌گردند. پردازش‌های نقطه‌ای شامل افزایش کنتراست، آستانه‌گیری سراسری، برابرسازی هیستوگرام و تطبیق هیستوگرام هستند و فیلترهای پایین‌گذر و بالاگذر، فیلتر میانه، فیلترهای ماکسیمم و مینیمم و آستانه‌گیری محلی از جمله پردازش‌های ماسک می‌باشند [۱۲].

حذف نویز: نویز ایجادشده به واسطه‌ی دستگاه‌های اسکنر نوری منجر به ایجاد نقطه نقطه‌های لک‌مانند^۴، قطعه خط‌های گسسته^۵، اتصال بین خطوط، فضاهای خالی در خطوط متن، پرشدن حفره‌های موجود در تصویر برخی حروف و ... می‌گردد. همچنین اعوجاج‌های مختلف شامل تغییرات محلی، منحنی‌شدن گوشه‌های حروف، تغییر شکل یا خوردگی حروف را نیز باید در نظر داشت. قبل از مرحله‌ی بازشناسی حروف، لازم است که این نقایص برطرف شوند. به منظور حذف این نویزها در تصاویر سطوح خاکستری از فیلترهای خطی و غیرخطی استفاده می‌شود [۱۳]. در این تصاویر، فیلترهای میانه و پایین‌گذر مانند فیلترهای میانگین یا تاری‌گوسین^۶ برای حذف نویز نقطه‌ای استفاده می‌گردند. همچنین، فیلترهای میانگین و گوسین برای حذف تاری نیز مفید هستند. نویزهای پربودیک که مانند قطار ضربه ظاهر شده و در طیف فوریه قابل مشاهده‌اند، با استفاده از فیلترهای notch مانند فیلتر باترورث، حذف می‌شوند. کاهش نویز، همچنین سبب کم‌شدن اندازه‌ی فایل تصویر می‌شود که به نوبه‌ی خود، کاهش زمان مورد نیاز برای پردازش‌ها و ذخیره‌سازی‌های آینده را در پی خواهد داشت [۱۴، ۱۵].

¹ Spatial filtering operations

² Point processing

³ Mask processing

⁴ Speckle

⁵ Disconnected line segments

⁶ Gaussian blur filter

اصلاح کجی تصویر^۱: از آنجایی که بسیاری از روش‌های آنالیز تصویر، به چرخش تصویر ورودی حساس هستند، نیاز است که پیش از دسته‌بندی، تصویر ورودی چرخانده شود. از جمله تکنیک‌های اصلاح کجی می‌توان به تبدیل هاف، مولفه‌های متصل و تکنیک‌های هم‌بستگی بین خطوط اشاره کرد. هال^۲ و تیلور^۳ بیست و پنج روش مختلف برای اصلاح کجی معرفی کرده‌اند که انتخاب یکی از این روش‌ها، به کاربرد و نوع تصاویری که استفاده می‌شوند، وابسته است [۱۶]. شکل ۱-۵، تصویر کج یک متن اسکن شده و پس از اصلاح کجی آن را نشان می‌دهد.

[illegible][illegible]

(ب)

(الف)

شکل ۱-۵: الف) تصویری از یک متن اسکن شده به صورت کج، ب) پس از اصلاح کجی آن.

قطعه‌بندی صفحه: بعد از بهبود تصویر، حذف نویز و اصلاح کجی، متن باید از تصاویر پشت‌زمینه و خطوط جدا شود. نتیجه‌ی این مرحله، تصویری تنها با متن است. تکنیک‌های بالا به پایین، پایین به بالا و ترکیبی، برای قطعه‌بندی صفحه به کار می‌روند.

نرمال سازی اندازه تصویر: بسته به کاربرد و روش های استخراج ویژگی یا تکنیک های کلاس بندی، تصاویر به یک اندازه ی مشخص، نرمال می شوند تا ویژگی ها از همگی کاراکترها در اندازه ی یکسان استخراج شوند و یکنواختی داده ها ایجاد گردد [۱۳].

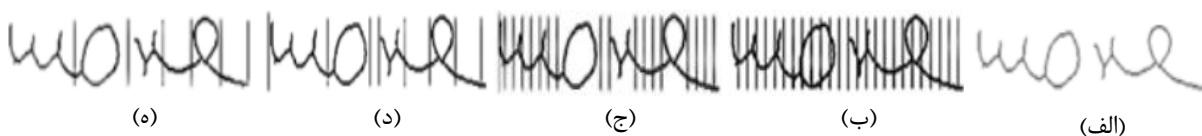
¹ Skew detection/correction² Hull

³ Taylor

۱-۶-۲- قطعه‌بندی

هدف این مرحله، قطعه‌بندی کردن خطوط متن تصویر به جملات، کلمات و حروف است [۱۷]. قطعه‌بندی یک مرحله‌ی مهم است؛ زیرا استخراج حروف در این مرحله، به طور مستقیم بر فرآیند بازشناسی تاثیر می‌گذارد. تکنیک‌های زیادی برای قطعه‌بندی کاراکتر توسعه داده شده‌اند و بیشتر آنها مخصوص متون خاصی هستند و بر روی تمام متون پاسخ نمی‌دهند. دو نوع اصلی قطعه‌بندی وجود دارد: صریح و غیر صریح.

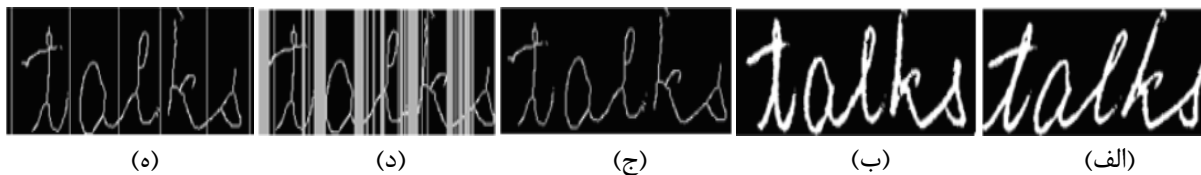
در روش غیر صریح، قطعه‌بندی و بازشناسی کاراکتر، همزمان صورت می‌گیرد. قطعه‌بندی به عنوان یک فرآیند جانبی ساده در بازشناسی است که تمام قطعه‌بندی‌های ممکن را فرض می‌کند و بازشناسی‌کننده، تصمیم می‌گیرد که بهترین فرض برای قطعه‌بندی کدام است. این روش، محاسبات بالایی دارد؛ زیرا تمام فرضیات ممکن را ایجاد کرده و تمام آنها برای انتخاب بهترین فرض در هنگام بازشناسی، ارزیابی می‌شوند. شکل ۶-۱، نحوه‌ی قطعه‌بندی غیر صریح را بر روی یک متن دست‌نوشته نشان می‌دهد. ابتدا تصویر به فاصله‌ی x ، که $x=h/18$ (ارتفاع تصویر کلمه و ۱۸ به صورت تجربی به دست آمده است)، قطعه‌بندی می‌شود (شکل ۶-۱ (ب)). برای هر خط عمودی، تعداد تقاطعات با پیکسل‌های پشت‌زمینه محاسبه می‌شود. اگر تعداد از یک بزرگتر بود، آن خط جداکننده، استخراج می‌شود (شکل ۶-۱ (ج)). اگر خطوط استخراجی عمودی در مرحله‌ی قبل در فاصله‌ی x هستند، اولین و آخرین خطوط عمودی، انتخاب می‌شوند (شکل ۶-۱ (د)). اگر خطوط جداکننده‌ی عمودی در فاصله‌ی کمتر یا مساوی x هستند، میانگین آنها به عنوان جداکننده‌ی بین حروف برگزیده می‌شود (شکل ۶-۱ (ه)).



شکل ۶-۱: قطعه‌بندی یک متن دست‌نوشته با استفاده از روش غیر صریح.

در روش صریح، تعداد پیکسل‌های پیش‌زمینه به صورت عمودی در هر ستون شمارش می‌شود تا فرورفتگی‌ها (گودی‌های حروفی مانند m) مشخص شوند. تنها ستون‌هایی به عنوان ستون‌های کاندید

قطعه‌بندی (CSC) ^۱ انتخاب می‌شوند که حاصل این شمارش برای آنها صفر یا یک باشد. در نتیجه، مشکل over-segmentation در دو حالت رخ می‌دهد. حالت اول این است که حروف به هم پیوسته، تعداد زیادی CSC خواهند داشت و حالت دوم برای حروفی مانند m, n, w و ... رخ می‌دهد که در حالت عمودی تعقیری ندارند. برای رفع این مشکلات، میانگینی از CSC‌هایی که فاصله‌ی آنها کمتر از یک حد آستانه است به یک ستون جداکننده ادغام می‌شوند. شکل ۷-۱، قطعه‌بندی یک متن دست‌نوشته به روش صریح را نشان می‌دهد [۱۸، ۱۹].



شکل ۷-۱: (الف) متن دست‌نوشته‌ی ورودی، (ب) تصویر ورودی بعد از رفع کجی، (ج) نازک‌سازی تصویر ورودی، (د) انتخاب بیش از چندین ستون کاندید قطعه‌بندی، (ه) میانگین‌گیری از ستون‌های کاندید و انتخاب تنها یک ستون به عنوان جداکننده و قطعه‌بندی نهایی تصویر ورودی.

۱-۶-۳- استخراج ویژگی‌ها

این مرحله، یکی از مراحل بسیار با اهمیت در سیستم‌های «بازشناسی نوری حروف» است؛ چرا که نتایج حاصل از این مرحله، مستقیماً بر روی کیفیت مرحله‌ی بازشناسی اثر می‌گذارد. در مرحله‌ی استخراج ویژگی‌ها، به هر الگوی ورودی، یک کد یا بردار ویژگی نسبت داده می‌شود که معرف آن الگو در فضای ویژگی‌ها است و آن را از دیگر الگوها متمایز می‌سازد [۹]. تکنیک‌های مختلف استخراج ویژگی به سه دسته‌ی اصلی تقسیم می‌شوند: ویژگی‌های ساختاری ^۲، ویژگی‌های آماری ^۳، و تبدیلات سراسری ^۴ [۲۰، ۲۱]. از ویژگی‌های ساختاری می‌توان به تعداد و نوع پاره‌خط‌های مختلف (عمودی، افقی، بالارونده و پایین‌رونده) اشاره کرد. گشتاورها و مکان‌های مشخصه، ناحیه‌بندی، نقاط تقاطع و فاصله از ویژگی‌های آماری هستند.

^۱ Candidate Segmentation Columns (CSC)

^۲ Structural features

^۳ Statistical features

^۴ Global transformation

۱-۶-۴- طبقه بندی و شناسایی

در آخرین مرحله‌ی فرآیند شناسایی کاراکتر، کاراکترها بر اساس ویژگی‌های استخراج شده، دسته‌بندی و شناسایی می‌شوند. رویکردهای اصلی در این مرحله به صورت زیر هستند:

۱. تطبیق الگوها
۲. تکنیک‌های آماری
۳. تکنیک‌های ساختاری

تطبیق الگوها: ساده‌ترین راه شناسایی کاراکتر، تطبیق نمونه‌های از پیش تعریف شده با کاراکتری است که باید شناسایی شود. یک کاراکتر با شباهتش به نمونه‌ها، شناسایی می‌شود. نرخ شناسایی این روش، به نویز و تغییر تصویر بسیار حساس است. فاصله‌ی اقلیدسی، معیارهای Jaccard و Mahalanobis و k نزدیکترین همسایه در این دسته قرار می‌گیرند.

تکنیک‌های آماری: توابع تصمیم آماری و تعدادی معیارهای بهینه‌سازی وجود دارند که برای شناسایی استفاده می‌شوند. کلاسه‌بند بیزین^۱، آنالیز دسته‌بندی، مدل مخفی مارکوف، مجموعه‌های فازی و کلاسه‌بند درجه دوم از جمله تکنیک‌های آماری هستند [۲۲].

تکنیک‌های ساختاری: این روش‌ها الگوهای ترکیبی را به منظور اندازه‌گیری شباهت بین مولفه‌های ساختاری، می‌سازند و هر کلاس قانون خاص خودش را برای توصیف ساختار همان کلاس دارد.

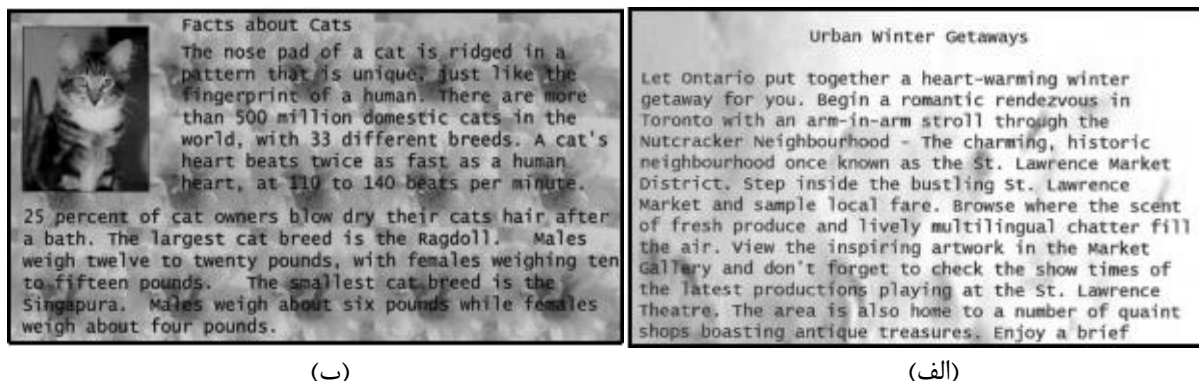
۱-۷- ارزیابی سیستم‌های بازشناسی نوری حروف

عملکرد سیستم‌های بازشناسی نوری بسته به تعدادی پارامتر، در متون مختلف، متفاوت است. برخی از این پارامترها عبارتند از: کیفیت کاغذ (کاغذهای با کیفیت خوب یا کاغذهای قدیمی)، فرآیند اسکن کردن متون، کیفیت اسکنر، رزولوشن اسکن، نوع اسناد چاپ شده (چاپگر لیزری یا فتوکپی)، نوع نوشتن متون (دست‌نوشته یا تایپی)، نوع فونت و اندازه. این پارامترها ممکن است به تارکردن^۲، نویزی کردن و رزولوشن پایین تصویر، منجر شوند و مصنوعاتی بر روی کاراکترهای اسکن شده ایجاد کنند [۲۳].

^۱ Bayesian

^۲ blur

هم‌چنین، روشنایی‌های غیریکنواخت در تصویر سند، پارامتر دیگری است که در دقت سیستم «ا» سی. آر» تاثیر می‌گذارد. برای مثال، شکل ۸-۱ (الف)، یک سند سطح خاکستری و (ب) تصویر سند با پس‌زمینه‌ی طرح‌دار را نشان می‌دهد که بازشناسی متن تصویر (الف) به دلیل ساده‌تر بودن ساختار پس‌زمینه، موفق‌تر از بازشناسی متن تصویر (ب) خواهد بود. عوامل دیگر شامل ویژگی‌های چاپ مانند یکنواختی، هم‌ترازی و چیدمان متن در صفحه، گرافیک و محتوای تصویر نیز موثر در عملکرد سیستم بازشناسی هستند [۱۲].



شکل ۸-۱: مثال‌هایی از تصاویر اسناد با پس‌زمینه‌های طرح‌دار و غیریکنواخت.

۸-۱- خلاصه و نتیجه‌گیری

در این فصل، به بررسی سیستم‌های بازشناسی نوری حروف پرداختیم. سیستم‌های گوناگونی برای شناسایی کاراکترها پیشنهاد شده است اما به دلیل عدم موفقیت ۱۰۰ درصدی در مرحله‌ی قطعه‌بندی و عدم استخراج مناسب‌ترین ویژگی‌ها برای تشخیص کاراکترها، هنوز هیچکدام از سیستم‌ها به دقت ۱۰۰ درصدی در شناسایی نرسیده‌اند. سیستم‌های بازشناسی نوری حروف عموماً شامل چهار مرحله هستند و الگوریتم‌های گوناگونی در هریک از این مراحل استفاده می‌شوند.

ارزیابی سیستم‌های بازشناسی نوری حروف بر اساس نرخ دقت در تشخیص کاراکترها، بررسی می‌شود و به عوامل گوناگونی مانند کیفیت کاغذ، فرآیند اسکن کردن، نوع نوشته، نوع فونت و ویژگی‌های چاپ وابسته است.

فصل دوم

مروری بر کارهای پیشین

۲-۱- مقدمه

در این فصل با آگاهی از مقدمات اولیه در مورد سیستم‌های بازشناسی نوری حروف و بیان ضرورت استفاده از آن، به توضیح خلاصه‌ای از تحقیقات صورت گرفته در این حوزه خواهیم پرداخت. از آنجایی که موضوع مورد بحث، سیستم‌های بازشناسی بدون محدودیت در اندازه و نوع فونت است، بررسی تحقیقاتی در این زمینه را در بخش سوم شرح می‌دهیم. نتیجه‌گیری، پایان‌بخش این فصل خواهد بود.

۲-۲- مروری بر روش‌های موجود بازشناسی نوری حروف

همانطور که در فصل اول گفته شد، بازشناسی نوری حروف برای شناسایی هر دو نوع متون چاپی و دست‌نوشته کاربرد دارد و همه ساله تحقیقات گسترده‌ای برای ارائه‌ی سیستم‌های بازشناسی نوری حروف با کارایی و سرعت بهتر انجام می‌شود.

شناسایی کاراکترهای دست‌نویس به دلیل تنوع دست‌خط و مدل‌های نوشتاری، دشوار است؛ زیرا سبب ایجاد اختلاف‌های بسیاری در کاراکترها می‌شود. صرف‌نظر از اینکه کدام کاراکتر در نظر گرفته شود؛ فرآیند استخراج ویژگی، یک نقش اساسی در شناسایی کاراکترها ایفا می‌کند. تاکنون هزاران ویژگی مختلف مانند ویژگی‌های مبتنی بر ضرایب حاصل از تبدیلات ریاضی، ویژگی‌های گشتاوری، ویژگی‌های مبتنی بر گراف، ویژگی‌های هندسی و ... در تحقیقات مختلفی مورد استفاده قرار گرفته‌اند [۲۴]. یکی از روش‌های استخراج اطلاعات مفید مشخصه‌های محلی الگوها، روش‌های zoning است که در شناسایی الگوهای چاپی نیز استفاده می‌شوند [۲۵]. در [۲۶]، برای شناسایی کاراکترهای دست‌نوشته Thai^۱ از یک شبکه 4×3 به منظور طراحی zoning مطابق با ویژگی‌های end-point و closed-loop از تصویر الگو، استفاده شده است. در [۲۷] کاراکترهای دست‌نویس از طریق یک zoning سلسله‌مراتبی مبتنی بر درخت شناسایی می‌شوند که برای بهینه‌سازی مرحله‌ی استخراج ویژگی، الگوریتم ژنتیک را به کار می‌برند. این شناسایی‌کننده، بر روی ویژگی‌های سلسله‌مراتبی جهت‌دهی، خمیدگی و مرکز ثقل عمل می‌کند. گره‌های این درخت، zone‌های گرهی والد خود را نشان می‌دهند و گرهی ریشه به تصویر با تمام zone‌ها، اشاره می‌کند. در [۲۸]، با کمک شبکه عصبی و شبکه‌هایی^۳ در اندازه‌های مختلف (۷×۷، ۵×۱۱، ۷×۱۱

^۱ زبان Thai که در کشور Thailand استفاده می‌شود.

^۲ Grid

^۳ Grid

و 9×13)، به شناسایی کاراکترهای دست‌نوشته و چاپی پرداخته شده است. نتایج نشان می‌دهند که دقت شناسایی کاراکتر در این سیستم به رزولوشن تصویر کاراکتر، وابسته است و تنوع دست‌خط‌های افراد گوناگون با شبکه عصبی قابل پوشش نیست.

روش‌های کلاس‌بندی بر اساس آموزش از روی نمونه‌ها، به طور گسترده‌ای برای شناسایی کاراکترها مورد استفاده قرار گرفته و بهبود چشم‌گیری در میزان دقت آنها، حاصل گشته است. روش‌های آماری، روش‌های مبتنی بر شبکه‌های عصبی مصنوعی و استفاده از ماشین بردار پشتیبان، نمونه‌هایی از تکنیک‌های به کار رفته در روش‌های موجود شناسایی کاراکترها می‌باشند [۲۹].

از آنجایی که شبکه‌های عصبی مصنوعی، توانایی زیادی در دسته‌بندی داده‌ها دارند، بدین منظور در سال‌های اخیر برای دسته‌بندی کاراکترها، بسیار استفاده شده‌اند. در مرجع [۳۰]، با استفاده از شبکه‌های عصبی چندلایه، یک سیستم شناسایی ارقام دست‌نویس و همچنین روشی برای استخراج مشخصه‌ها بر اساس شکل رقم، معرفی شده است که دقت دسته‌بندی ارقام بر روی پایگاه داده MNIST با تعداد ۶۰۰۰۰ نمونه‌ی آموزشی و تعداد ۱۰۰۰۰ نمونه‌ی آزمایشی، به ۸۰٪ رسیده است. این پایگاه داده، حاوی تصاویر ارقام دست‌نوشته‌ی ۲۵۰ فرد مختلف می‌باشد. کارایی الگوریتم‌های تکاملی برای شبکه عصبی رو به جلو، با سه تکنیک مختلف محاسبات نرم^۱، به منظور شناسایی الفبای انگلیسی دست‌نویس در [۳۱] بررسی شده است. بررسی‌ها نشان می‌دهند که شبکه عصبی رو به جلو با دو الگوریتم تکاملی، دقت عموم‌گرایی بهتری در مسائل شناسایی کاراکتر، ایجاد کرده و الگوریتم تکاملی ترکیبی به طور موثرتر و قابل اطمینان‌تری، مساله را حل می‌کند. همچنین، این الگوریتم، پیچیدگی زمانی و فضایی کمتری برای داده‌های آموزشی دارد. در مرجع [۳۲]، سه سیستم مدل زبانی^۲ شبکه عصبی یک‌پارچه برای شناسایی کاراکترهای دست‌نوشته‌ی پایگاه داده‌ی IAM^۳، آزمایش شده‌اند که عبارتند از: شبکه‌های عصبی بازگشتی دوطرفه، مدل‌های مخفی مارکوف و ترکیبی از هر دو. نتایج آزمایش‌ها نشان می‌دهند که کاهش میزان خطای کلمه را می‌توان با مدل‌های زبانی شبکه عصبی در مقایسه با مدل‌های زبانی آماری N-gram به دست آورد. همچنین، بهترین نرخ خطای به دست آمده برای این پایگاه داده برابر ۱۶٫۱٪ با استفاده از مدل‌های زبانی شبکه‌ی عصبی است.

^۱ Soft-computing

^۲ Language model: یک احتمال را بر اساس تابع توزیع احتمالاتی، به یک دنباله از m کلمه اختصاص می‌دهد.

^۳ این پایگاه داده شامل متون انگلیسی دست‌نوشته‌ی ۶۵۷ نفر است.

در [۳۳] نرخ موفقیت شبکه‌ی عصبی هاپفیلد، در یادآوری کاراکترهای انگلیسی چاپی ذخیره شده، ارزیابی شده است. در این فرآیند، از الگوریتم ژنتیک برای تعیین وزن‌های بهینه‌ی یادآوری صحیح، متناظر با نمونه‌ی نویزی کاراکترها استفاده می‌شود. نتایج شبیه‌سازی، کارایی بهتر شبکه را برای یادآوری حروف ذخیره شده با استفاده از الگوریتم ژنتیک بر روی ماتریس وزن نیمه-بهینه، نشان می‌دهد. روش ارزیابی شباهت مکمل (CSM)^۱ که یک روش شناخته شده‌ی کلاس‌بندی است، در [۳۴] به کار برده می‌شود. هم‌چنین، CSM، علاوه بر دسته‌بندی کننده، یک استخراج‌کننده‌ی ویژگی نیز می‌باشد. بردار ویژگی CSM برای آموزش به یک شبکه عصبی چندلایه، داده می‌شود و سبب می‌گردد که شبکه‌ی عصبی تبدیل به یک ابزار بسیار قدرتمند و مناسب برای شناسایی کاراکترها شود. نتایج آزمایش‌ها، توانایی سیستم پیشنهادی را برای شناسایی کاراکترهای چاپی انگلیسی با کیفیت پایین نشان می‌دهند.

یک سیستم شناسایی کاراکترهای دست‌نوشته و چاپی با استفاده از مورفولوژی ریاضی در مرجع [۳۵] پیشنهاد شده است. ابتدا، مجموعه‌ای از ویژگی‌های ساختاری با اهمیت برای جداسازی کاراکترها از یکدیگر، استخراج می‌شوند. سپس، با استفاده از یک درخت تصمیم که در هر گره، قوانین تصمیم توسط عملگرهای مورفولوژی تعریف می‌گردند، کلاس‌بندی انجام می‌شود. دقت دسته‌بندی این سیستم با فرض هم‌پوشان نبودن کاراکترها، برای متون دست‌نوشته حدود ۹۵٪ و متون چاپی ۹۹٪ است. یک سیستم خودکار شناسایی متون چاپی و دست‌نوشته‌ی لاتین و عربی در [۳۶] ارائه شده است. این روش، بر اساس تحلیل سراسری بافت با استخراج ویژگی‌های چندبعدی فرکتال، عمل می‌کند. سیستم ارائه شده برای ۱۰۰۰ نمونه (۴۰۰ نمونه‌ی آموزشی، ۶۰۰ نمونه‌ی آزمایشی) با اندازه و سه نوع فونت متفاوت آزمایش شده و دقت آن با استفاده از KNN، ۹۶٫۶۴٪ و RBF، ۹۸٫۷۲٪ به دست آمده است.

تاکنون شناسایی کاراکترها به طور گسترده‌ای در زبان‌های مختلف، مطالعه شده است. مساله‌ی شناسایی کاراکترهای فارسی و عربی به سبب ویژگی‌های خاص مانند شکل آنها، پیوستگی، نقطه‌ها و فونت‌های متفاوت، یکی از مسائل به‌روز و محبوب به شمار می‌آید. تکنیک پیشنهادی در [۳۷]، از شبکه‌ی عصبی RBF برای شناسایی کاراکترهای عربی در یک فونت مشخص با اندازه‌های مختلف، استفاده می‌کند. پس از پیش‌پردازش متن و قطعه‌بندی آن به حروف، ویژگی‌های گشتاورهای مرکزی، استخراج شده و به‌وسیله‌ی شبکه‌ی RBF، دسته‌بندی می‌شوند. مرجع [۳۸]، یک سیستم شناسایی متن عربی

^۱ Complementary similarity measure (CSM)

پیشنهاد می‌دهد. این سیستم، تصویر متن را به تصاویر خطوط متن تجزیه کرده و مجموعه‌ای از ویژگی‌های آماری ساده را، از پنجره‌ای به طول یک که بر روی تصویر در حال گذر است؛ استخراج می‌کند. سپس بردارهای ویژگی برای شناسایی به مدل مخفی مارکوف، داده می‌شوند. سیستم پیشنهادی با بیش از ۶۰۰ متن عربی در اندازه‌ی برگه A4 در چندین فونت، ارزیابی می‌شود. یک سیستم شناسایی کاراکترهای عربی در چندین فونت و اندازه در [۳۹] معرفی شده است. این سیستم بر اساس مدل مخفی مارکوف تولکیت (HTK) و برنولی-مدل مخفی مارکوفی که در آن تابع چگالی احتمالی برنولی، جایگزین تابع چگالی گوسین، می‌شود - می‌باشد. چندین تست برای ارزیابی این سیستم بر روی پایگاه داده APTI، انجام شده است. دقت ۹۸,۳٪ برای شناسایی کاراکترها با تست یک فونت در اندازه‌های ۶، ۸، ۱۲، ۱۸ و ۲۴، حاصل شده است. با آزمایش پنج فونت، این دقت به ۱۰-۲۰٪ کاهش یافته است. در [۴۰] یک سیستم جدید برای شناسایی سریع کاراکترهای چاپی فارسی با استخراج ویژگی‌های گشتاورهای موجک زیرنیک^۱، پیشنهاد شده است که شناسایی آن‌ها با شبکه‌های عصبی مصنوعی، صورت می‌گیرد. نتایج شبیه‌سازی نشان می‌دهند که دقت شناسایی این سیستم به طور میانگین ۴,۳۷ بار و سرعت آن، حدود ۸ بار بیشتر از کارهای مشابه است.

در [۴۱]، یک روش جدید برای شناسایی کاراکترهای اصلی Kannada با استفاده از روش zoning و ویژگی‌های آماری، پیشنهاد شده است. برای مقایسه‌ی نمونه‌های آموزشی و تست از دسته‌بند نزدیک‌ترین همسایه، استفاده می‌شود. این روش نسبت به نویز، تارشدگی و تفاوت در نوع و اندازه‌ی فونت، مقاوم بوده و پس از ارزیابی ۱۰۴۳ نمونه به دقت ۸۳,۴۹٪ رسیده است. یک سیستم شناسایی کاراکترهای انگلیسی و Kannada در [۴۲] پیشنهاد شده است. ۲۸۰۰ کاراکتر Kannada و ۲۳۰۰ کاراکتر کوچک انگلیسی پس از نرمال شدن به اندازه ۳۲×۳۲ توسط ماشین بردار پشتیبان 2-fold cross-validation دسته‌بندی شدند. هر تصویر نرمال شده به ۶۴ zone ناهم‌پوشان، تقسیم و چگالی پیکسل برای هر zone، محاسبه شده است. میانگین دقت به دست آمده برای شناسایی کاراکترهای Kannada و انگلیسی به ترتیب ۷۳,۳۳٪ و ۹۶,۱۳٪ است. مرجع [۴۳]، یک سیستم شناسایی کاراکتر تاریخی ژاپنی با استفاده از ویژگی‌های جهت‌دار^۲ و شبکه‌های عصبی ماژولار، ارائه می‌دهد. شبکه‌های عصبی ماژولار شامل دو نوع شبکه عصبی نگاشت خودسازمانده و شبکه عصبی چندلایه، می‌شود که هریک از آنها یک مساله‌ی دو

¹ Zernike

² Directional element features

کلاسه را کلاس‌بندی می‌کنند. نتایج حاصل، موثرتر بودن سیستم پیشنهادی را در مقایسه با تحقیقات گذشته در دسته‌بندی ۵۷ دسته از حروف نشان می‌دهند. در مرجع [۴۴]، ارقام دست‌نویس هندی ۱ تا ۹، با استفاده از یک شبکه‌ی عصبی احتمالاتی، بازشناسی می‌شوند. بردار ویژگی بر اساس مرکز جرم و مجموعه‌ای از بردارهای نقاط مرزی رقم می‌باشد که مستقل از مقیاس، چرخش و انتقال است. دقت حاصل از شناسایی ارقام با بردار ویژگی به طول ۱۲، ۹۹٫۷۲٪ می‌باشد. مرجع [۴۵]، روشی برای شناسایی کاراکترهای سریانی در متون تاریخی، ارائه می‌دهد. بعد از قطعه‌بندی متون توسط روش آماری جدید، ویژگی‌هایی متناسب با شکل کاراکترها، استخراج و با استفاده از ماشین بردار پشتیبان با 10 fold cross-validation، بازشناسی می‌شوند. آزمایش‌هایی با چندین روش قطعه‌بندی و ترکیبی از ویژگی‌های مختلف، بر روی متون باقی‌مانده از قرن نوزدهم، صورت گرفته و بسته به الگوریتم‌های مورد استفاده، اندازه و منبع متن، دقت ۶۱٪ تا ۱۰۰٪ حاصل شده است. با استفاده از ویژگی‌های ساختاری و آماری، یک سیستم شناسایی کاراکترهای چینی در [۴۶] پیشنهاد شده است که ویژگی آماری، کلاس اولیه را تعیین و ویژگی ساختاری، کاراکترهای چینی را بازشناسی می‌کند و یک الگوریتم جدید بر اساس دسته‌بند فاصله‌ی مینیمم، پیشنهاد می‌شود. نتایج آزمایش‌ها، موثر بودن الگوریتم شناسایی را برای کاراکترهای چینی، اثبات می‌کند.

یکی از اساسی‌ترین مراحل در تحلیل اسناد و تشخیص متون، بازشناسی قلم است که شامل تعیین مواردی چون زبان متن، اندازه قلم و نوع قلم می‌باشد. بازشناسی قلم قبل از مرحله‌ی بازشناسی حروف، می‌تواند کارایی سیستم درک سند را به مراتب بالا ببرد چرا که با استفاده از اطلاعات قلم، می‌توان پارامترهای سیستم را به طور خاصی تنظیم کرد. طبق گزارشات منتشر شده، سیستم‌های بازشناسی حروف که برای متون تک قلمی طراحی می‌شود نرخ بازشناسی بالاتری نسبت به سیستم‌هایی دارند که برای متون چند قلمی طراحی شده و بازشناسی را مستقل از نوع قلم انجام می‌دهند.

در ترکیب یک سیستم بازشناسی نوری حروف و بازشناسی نوری قلم، دو نوع رویکرد وجود دارد [۴۷]: بازشناسی پیشین قلم که قلم بدون اطلاع از حروف بازشناسی می‌شود و بازشناسی پسین قلم که از اطلاعات حروف برای بازشناسی قلم استفاده می‌کند و در مقاصد چاپی کاربرد دارد. روش‌های ارائه شده برای بازشناسی قلم در سطوح مختلفی ارایه می‌شود: روش‌های مبتنی بر تصویر

کل متن، روش‌های مبتنی بر اطلاعات یک خط متن، روش‌هایی که کلمات متن را بررسی می‌کند و روش‌هایی که از اطلاعات هر حرف به صورت مجزا استفاده می‌کنند. در هر کدام از روش‌های ذکر شده، نوع قلم بر اساس ویژگی‌هایی از کل متن، یک خط متن، هر کلمه به صورت مجزا یا هر حرف استخراج شده، تشخیص داده می‌شود. بر اساس دانش قبلی ما که در چه سطحی از متن، نوع قلم یکسان است، سیستم در سطح مناسب طراحی می‌شود؛ ضمن اینکه در سطوح بالاتر سرعت پردازش بالاتر است [۴۸].

در مرجع [۴۹] از ویژگی‌های نوشتاری برای تشخیص ۷ نوع قلم متداول انگلیسی استفاده شده است. هر خط تصویر به سه ناحیه صعودی^۱، ارتفاعی^۲ و نزولی^۳ تقسیم و برای هر نوع قلم، انواع سریف‌ها^۴ در نواحی مختلف استخراج و بازشناسی به کمک شبکه عصبی پرسپترون چند لایه با دقت ۹۵٪ انجام شده است. در [۵۰] برای طبقه بندی ۶ نوع قلم چینی و ۸ نوع قلم انگلیسی، تمام متن به صورت یک تصویر بافت‌دار در نظر گرفته شده است. برای استخراج ویژگی از ۱۶ کانال فیلتر گابور در ۴ جهت و ۴ مقیاس مختلف استفاده شده و میانگین و انحراف استاندارد تصویر خروجی هر کانال به عنوان ویژگی مناسب استخراج شده است. در مرجع [۵۱] برای طبقه بندی ۲۰ نوع قلم متداول فارسی، از ۲۴ فیلتر گابور در ۴ مقیاس و ۶ جهت استفاده شده است. میانگین و انحراف معیار تصویر خروجی هر فیلتر به عنوان ویژگی در نظر گرفته شده است. یک سیستم جدید برای شناسایی فونت‌های فارسی با ویژگی‌های مبتنی بر گرادیان‌های روبرتز^۵ و سوبل^۶ (SRF) در ۱۶ جهت در [۵۲] ارائه شده است. پس از تقسیم یک خط از متن به چندین بخش کوچک، SRF به عنوان ویژگی استخراج می‌شود. این ویژگی، زمان محاسبات کمتری دارد و خیلی سریع‌تر از ویژگی‌های دیگر مانند فیلتر گابور و ویژگی‌های گشتاوری استخراج می‌شود. دقت شناسایی این سیستم، ۹۴٫۱۶٪ برای ده فونت محبوب فارسی است. در [۴۸]، از ضرایب خود همبستگی به عنوان ویژگی‌های مناسب برای بازشناسی قلم‌های فارسی استفاده می‌شود. ایده‌ی مطرح شده این است که مجموعه‌ای از ضرایب فوق، حاوی اطلاعات مفیدی از فرم و شکل قلم می‌باشند که می‌توان از آن‌ها جهت طبقه‌بندی استفاده کرد. همچنین، برای

^۱ Ascender zone

^۲ Height zone

^۳ Descended zone

^۴ خط کوچکی که به بالا یا پایین کاراکترها در برخی از فونت‌ها اضافه می‌شود.

^۵ Roberts

^۶ Sobel

کاهش بعد ویژگی‌های پیشنهادی، الگوریتم وراثتی به کار گرفته شده است. روش فوق بر روی مجموعه‌ای شامل ۱۰۰۰ زیر تصویر از ۲۰ نوع قلم متداول اعمال و نرخ بازشناسی ۹۶,۹٪ با استفاده از طبقه بند k همسایه نزدیکتر حاصل شده که مناسب و کارا بودن روش پیشنهاد شده را می‌رساند. در مرجع [۵۳]، از ویژگی‌های گشتاورهای مرتبه سوم و چهارم، برای طبقه بندی ۸ نوع قلم اسپانیایی در ۴ حالت مختلف استفاده شده است. در این سیستم، تصویر خاکستری متن در یک بردار یک بعدی چیده شده، ۱۶ گشتاور مرتبه سوم و چهارم در ۴ جهت ۴۵، ۰، ۹۰ و ۱۳۵ درجه، از هر تصویر استخراج شده است.

۲-۳- بازشناسی نوری حروف بدون محدودیت در اندازه و نوع فونت

یکی دیگر از زمینه‌های کاربردی شناسایی کاراکترها، تشخیص آنها بدون محدودیت در اندازه و نوع فونت است. اگر سیستم بازشناسی، موجود باشد که بتواند کاراکترها را مستقل از نوع فونت، تشخیص دهد؛ دیگر نیازی به تغییر سیستم‌های بازشناسی قبلاً معرفی شده، برای شناسایی کاراکترهای با فونت‌های جدید نیست.

کارهای معدودی برای شناسایی ارقام و کاراکترها در چندین فونت و اندازه در زبان‌های گوناگون انجام شده؛ از جمله سیستمی برای بازشناسی حروف و ارقام تایپی Kannada بدون محدودیت در اندازه و فونت، با استفاده از گشتاورهای ثابت در [۵۴] پیشنهاد شده است. از دسته‌بند نزدیک‌ترین همسایه با فاصله‌ی مینیمم، برای دسته‌بندی استفاده می‌شود. این روش، بر روی ۱۸۰۰ تصویر از حروف و ۱۰۰۰ تصویر از ارقام آزمایش شده و دقت ۹۷,۷٪ از دسته‌بندی حروف و ۹۸,۹۲٪ از دسته‌بندی ارقام به دست آمده است. در مرجع [۵۵]، روشی برای شناسایی ارقام فارسی چاپی، پیشنهاد داده شده است. در این روش، توصیف‌کننده‌های تصویر رقم، گشتاورهای مرکزی هندسی هستند و ارقام با استفاده از شبکه‌ی عصبی فازی min-max دسته‌بندی می‌شوند. بر اساس گزارشی که در این مقاله آمده؛ روش ارائه شده با دقت ۹۹,۱۶٪ می‌تواند داده‌های آموزش یافته را حتی با اعمال تغییراتی چون تغییر اندازه، تبدیل و یا چرخش، شناسایی نماید. یک روش ترکیبی برای شناسایی ارقام فارسی در چندین فونت به وسیله‌ی [۵۶] پیشنهاد شده است. پس از محاسبه‌ی میانگین زاویه و فاصله‌ی هر یک از نواحی به دست آمده از تقسیم تصویر به عنوان بردار ویژگی، برای تشخیص ارقام از مجموعه‌های فازی، استفاده شده است. بعد از انجام

عملیات دِفازی کردن^۱، سه کاندید از ارقام انتخاب می‌شوند که این کاندیداها با ویژگی‌های دیگری در یک سیستم فازی جدید، پس‌پردازش می‌گردند که با پایگاه داده‌ی ۲۱۶ تایی در ۲۴ فونت، دقت ۹۷٫۵٪ حاصل شده است. در مرجع [۵۷]، یک موتور استنتاج نئورو-فازی برای بازشناسی ارقام فارسی، ارائه شده است. در این سیستم، از موتور استنتاج فازی ممدانی^۲ برای قوانین فازی بهره گرفته شده که با آموزش شبکه‌ی عصبی پرسپترون روی ویژگی‌های ۳۳ نوع فونت مختلف، به دقت ۹۷٪ بدون محاسبه‌ی رقم چهار رسیده‌اند؛ با حضور رقم چهار، دقت سیستم بیشتر از ۹۵٪ نیست که تأکید می‌کند سیستم پیشنهادی، نیاز به تصحیح دارد.

یک روش جدید که دقت خوبی در بازشناسی کاراکترهای با سه فونت مختلف و تخمین زاویه‌ی آن‌ها دارد، در [۵۸] ارائه شده است. در این مرجع، با استفاده از ماتریس کواریانس تعدادی از الگوهای دارای چرخش، یک زیر-فضای eigen^۳ برای هر دسته ایجاد می‌شود. سپس یک مکان هندسی، توسط نگاشت کاراکترهای دارای چرخش به زیرفضای eigen به دست می‌آید. با محاسبه‌ی فاصله‌ی نقاط نگاشت یافته از یک کاراکتر ناشناخته و مکان هندسی به دست آمده از آن، کاراکتر، شناسایی می‌شود. در مرجع [۵۹]، با افزودن یک ویژگی جدید بر اساس مشخصه‌ی تناوبی نقاط نگاشت یافته بر روی فضای eigen به سیستم پیشنهادی مرجع [۵۸]، علاوه بر شناسایی حروف انگلیسی بزرگ، سیستم قادر است کاراکترهای کوچک و ارقام انگلیسی را بازشناسی کند. این دو سیستم پیشنهادی، از روش استخراج ویژگی پیچیده‌ای برخوردارند؛ هم‌چنین، محاسبه‌ی ماتریس کواریانس و استخراج زیرفضایی از مقادیر ویژه، عملی زمان‌بر و دارای حجم بالایی از محاسبات است. تشخیص تنها سه فونت از تعداد زیاد فونت‌های موجود، کاربرد بسیار کم سیستم‌های [۵۷] و [۵۸] را در دنیای واقعی نشان می‌دهد. ما به دنبال ارائه‌ی سیستمی هستیم که علاوه بر سادگی روش استخراج ویژگی، قادر به تشخیص تعداد زیادی از فونت‌ها باشد و در صورت افزایش تعداد فونت‌ها، دقت تشخیص به طور قابل ملاحظه‌ای کاهش نیابد.

شناسایی ارقام در ۱۶ فونت مختلف و اندازه‌های گوناگون در بازه‌ی ۸ تا ۷۲، توسط سیستم پیشنهادی [۶۰] صورت می‌گیرد. دقت بازشناسی ارقام پایگاه داده‌ی ۳۲۰۰ تایی برابر ۹۹٫۷۸٪ است. در این سیستم، با یک نگرش مستقل از نازک‌سازی، چگالی‌های بیرونی پیکسل‌ها برای هریک از چهار جهت

^۱ defuzification

^۲ Mamdani

^۳ Eigen Sub-space: زیرفضای انتخاب شده از بردار مقادیر ویژه (eigen values).

بالا، پایین، چپ و راست محاسبه می‌شود. سپس، یک درخت تصمیم با دسته‌بند نزدیک‌ترین همسایه، ارقام را دسته‌بندی می‌کند. یک سیستم شناسایی حروف انگلیسی در ۷ نوع فونت، بر اساس گشتاورهای ثابت در [۶۱] پیشنهاد شده است که ۳۶۴ نمونه برای آموزش استفاده می‌شوند. تصاویر پس از نرمال شدن به اندازه 33×33 و نازک‌سازی، با گشتاورهای ثابت معرفی شده توسط پالانیپان^۱ ارزیابی و یک درخت تصمیم بر اساس کلاسه‌بند نزدیک‌ترین همسایه، آنها را دسته‌بندی می‌کند که با تست ۷۲۸۰ نمونه، دقت ۹۹٫۶۵٪ حاصل شده است. مرجع [۶۲] یک سیستم برای شناسایی اسناد انگلیسی و Gurumukhi با ۱۷ فونت در ۱۱ اندازه‌ی متفاوت، ارائه می‌دهد. با استخراج بردار ویژگی ۱۸۹ تایی از ویژگی‌های گابور بر اساس فرکانس جهت‌دار و بردار ۲۰۰ تایی از ویژگی‌های گرادیان بر اساس اطلاعات گرادییانی هر کاراکتر، یک مساله‌ی ۴ کلاسه با ماشین بردار پشتیبان چندکلاسه، کلاس‌بندی می‌شود. نرخ میانگین دقت به دست آمده از شناسایی ۴۸۶۲ نمونه از حروف با آموزش نیمی از نمونه‌ها، برای ویژگی‌های گابور و گرادیان به ترتیب ۹۶٫۴۷٪ و ۹۸٫۰۸٪ است. تعداد زیاد ویژگی‌های استخراجی در روش معرفی شده در [۶۱] و استفاده از نیمی از نمونه‌ها برای آموزش، سبب افزایش حجم محاسبات می‌گردد. همچنین، به دلیل آموزش بسیاری از فونت‌ها به سیستم، این احتمال وجود دارد که در صورت حضور یک فونت جدید در نمونه‌های تست، سیستم قادر به تشخیص آن فونت نبوده و خطای تشخیص افزایش یابد. با آموزش تنها یک فونت از نمونه‌ها به سیستم و استفاده از مفهوم شباهت بین نمونه‌ها، می‌توان این اطمینان خاطر را ایجاد کرد که در صورت تست فونت جدیدی، سیستم قادر به تشخیص آن خواهد بود و دیگر نیازی نیست که آن فونت، قبلاً به سیستم آموزش داده شده باشد. سیستم پیشنهادی در این تحقیق، از این ایده جهت آموزش نمونه‌ها بهره می‌برد.

۴-۲- نتیجه‌گیری

در این فصل، تحقیقات انجام گرفته در حوزه بازشناسی متون دست‌نوشته و چاپی را شرح دادیم. در هر یک از این تحقیقات، روش‌های مختلف استخراج ویژگی و کلاس‌بندی استفاده شده‌اند که هر کدام مزایا و معایب مختص به خود را دارند. اکثر تحقیقات، سعی بر آن داشتند که بتوانند سیستمی را معرفی کنند که علاوه بر دقت بالای بازشناسی، سرعت بازشناسی خوبی نیز داشته باشند. در پایان نیز به شرح

¹ Palaniapan

سیستم‌های بازشناسی بدون محدودیت در اندازه و نوع فونت پرداختیم. در فصل بعدی، پیاده‌سازی سیستم بازشناسی پیشنهادی بدون محدودیت در اندازه و نوع فونت را توضیح خواهیم داد.

فصل سوم

پیاده‌سازی سیستم پیشنهادی

۳-۱- مقدمه

در این فصل، به شرح مراحل مختلف سیستم پیشنهادی می‌پردازیم. همانطور که در فصل اول بیان شد، یک سیستم بازشناسی نوری کاراکتر معمولاً به چهار مرحله تقسیم می‌شود: پیش‌پردازش، قطعه‌بندی، استخراج ویژگی‌ها و طبقه‌بندی و شناسایی. در سیستم پیشنهادی، مرحله‌ی قطعه‌بندی را نخواهیم داشت؛ زیرا تصاویر کاراکترها به صورت جداگانه تهیه شده و تاکنون روش‌های بسیار خوبی جهت انجام قطعه‌بندی معرفی شده‌اند [۶۳-۶۵] که با استفاده از آن روش‌ها، می‌توان با دقت بالایی، متون موردنظر را به کاراکترها قطعه‌بندی کرد. در ادامه، هرکدام از این مراحل را برای شناسایی ارقام و حروف انگلیسی توضیح خواهیم داد.

۳-۲- سیستم شناسایی پیشنهادی

۳-۲-۱- پایگاه داده

با توجه به آنکه پایگاه داده‌ی اولیه‌ی آماده‌ای در انواع فونت‌ها و اندازه‌های مختلف وجود ندارد، دو پایگاه داده برای شناسایی ارقام و حروف تولید کرده‌ایم. پایگاه داده به کار رفته برای شناسایی ارقام و حروف، تصاویر دودویی ارقام و حروف بزرگ و کوچک انگلیسی با فونت‌ها و اندازه‌های متفاوت در رزولوشن ۳۰۰ نقطه بر اینچ است. ابتدا، به دلیل آنکه بتوان سیستم پیشنهادی را با سایر کارهای انجام شده مقایسه کرد، پایگاه داده‌ای در فونت‌ها و اندازه‌های مشابه با آنها ایجاد کردیم. بدین منظور، ۱۸۷ نمونه از هر یک از ارقام و حروف بزرگ و کوچک در هفده فونت Times, Times new roman, Arial, Calibri, Cambria, ArialRoundedMTBold, Georgia, MicrosoftSansSerif, Comic, CenturySchoolBook, Garamond, Verdana, Coolvetica, CourierPS, Palatino, Bookman و NewCenturySchoolBook در یازده اندازه‌ی مختلف ۱۰، ۱۱، ۱۲ تا ۲۸ (اندازه‌های زوج) ایجاد کردیم؛ بنابراین، پایگاه داده‌ی ارقام شامل ۱۸۷۰ نمونه و پایگاه داده‌ی حروف شامل ۹۷۲۴ (۴۸۶۲ برای هریک از حروف بزرگ و کوچک) است. به منظور نشان دادن مقاومت سیستم پیشنهادی در برابر تغییر فونت، ۱۳ نمونه‌ی دیگر از هر یک از ارقام در سیزده فونت جدید Avant Garde, Courier, Lucida Bright, Lucida Console, Estrangelo Edessa, Sans, Book Antiqua, DejaVu sans serif, Trebuchet MS, Lucida Console, Estrangelo Edessa, Century gothic, Sylfean, Shruthi and Tahoma و اندازه‌ی ۲۰ به پایگاه داده‌ی ارقام افزودیم که در

مجموع تعداد نمونه‌های این پایگاه داده به ۲۰۰۰ می‌رسد. در پیوست ۲، نمونه فونت‌های مختلف هر یک از ارقام در پایگاه داده آمده است.

به منظور شناسایی فونت‌های بیشتر حروف، ۳۵ نمونه از هر یک از حروف در هفت فونت Lucida Felix و sans Typewriter, Rockwell, Bodoni MT, Tahoma, Harry P, Copperplate gothic light Titling و اندازه‌های ۱۶، ۱۸، ۲۰، ۲۲ و ۲۴ به صورت نرمال را نیز به پایگاه داده افزودیم. ۱۵ نمونه از هریک از حروف نیز به صورت پررنگ^۱ و زاویه‌دار^۲ با ۸ فونت Times New Roman, Arial, Bodoni MT, Calibri, Cambria, Tahoma^۳, Rockwell, Century gothic مجموع تعداد نمونه‌های پایگاه داده‌ی حروف به ۱۲۳۲۴ (۶۱۶۲ نمونه برای هریک از حروف بزرگ و کوچک) می‌رسد. پیوست ۲، فونت‌های مختلف موجود در پایگاه داده را برای حروف بزرگ و کوچک و نمونه‌های پررنگ و زاویه‌دار آنها را نشان می‌دهد. چون هدف ارائه‌ی این سیستم، مقایسه‌ی نمونه‌ها بر اساس شباهت بین آنهاست، بنابراین، با آموزش تعداد کمی از نمونه از هریک از حروف و ارقام، سیستم باید قادر باشد سایر نمادهای مشابه با آنها را تشخیص دهد؛ در نتیجه، برای تهیه‌ی مجموعه نمونه‌های آموزشی ارقام و حروف، از هر رقم تنها یک نمونه به صورت تصادفی و دو نمونه از هر یک از حروف انتخاب شده است.

چون تصاویر، سیاه و سفید یا به عبارتی دیگر باینری هستند؛ پیکسل‌های تصاویر، تنها مقادیر صفر و یک را دارند. مقادیر صفر، نواحی سفیدرنگ و مقادیر یک، نواحی سیاه را نشان می‌دهند. بنابراین در یک تصویر، پیکسل‌های کاراکتر تنها شامل یک‌ها و پس‌زمینه‌ی آن تنها شامل صفرها هستند.

۳-۲-۲- پیش‌پردازش

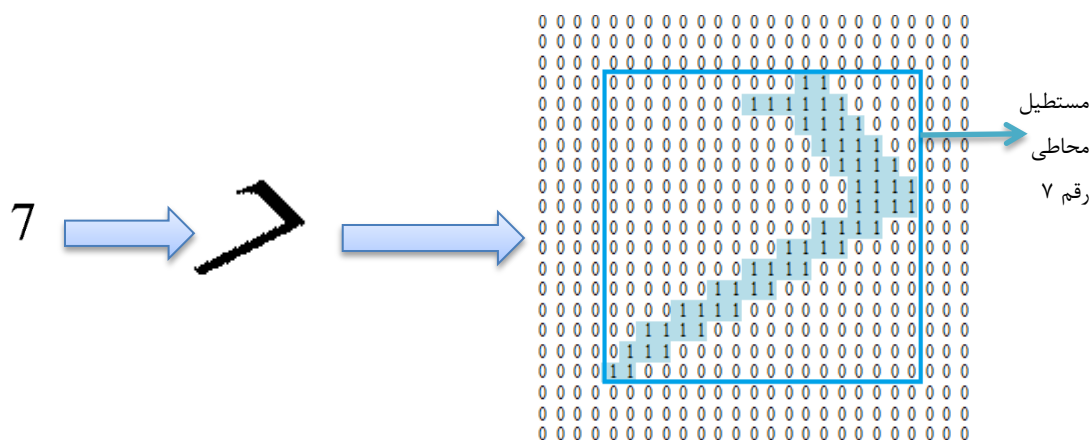
پیش از تشکیل بردار مشخصه، لازم است پیش‌پردازش‌هایی بر روی تصویر ورودی صورت گیرد. معمولاً هنگامی که کاراکترها اسکن می‌شوند با نویز همراه می‌باشند که با فیلتر میانه می‌توان آن را به خوبی برطرف کرد. البته نمونه‌هایی که برای تست و آزمایش روش پیشنهادی در این تحقیق استفاده کرده‌ایم فاقد نویز هستند. پس از رفع نویز، تصویر را به اندازه‌ی ۴۵ درجه در جهت خلاف عقربه‌های ساعت می‌چرخانیم. علت این امر در بخش ۳-۲-۳، توضیح داده خواهد شد. سپس، برای تصویر هر کاراکتر

^۱ Bold

^۲ Italic

^۳ فقط نمونه‌های پر رنگ فونت Tahoma در پایگاه داده موجود است.

با استخراج کوچکترین مستطیل محاط بر کاراکتر مربوطه، خود کاراکتر را از پس‌زمینه جدا می‌کنیم. از آنجایی که فونت و اندازه‌ی کاراکترهای مختلف یکسان نیستند؛ سایز مستطیل‌های محاطی آن‌ها نیز یکسان نخواهند بود. بنابراین برای نرمال‌سازی اندازه‌ی تصاویر، مشابه کار [۶۶] تصاویر را به اندازه‌ی 30×30 پیکسل تغییر می‌دهیم. تغییر اندازه‌ی تصویر توسط روش درونیابی نزدیکترین همسایه صورت می‌گیرد. این روش، ساده و بدون محاسبات پیچیده‌ی ریاضی است و برای داده‌های گسسته مناسب می‌باشد. به علاوه این روش نسبت به سایر روش‌های درونیابی مانند روش‌های خطی [۶۷]، درجه دو^۱ [۶۸] و درجه سوم^۲ [۶۹] از سرعت بسیار زیادی برخوردار است [۷۰]. شکل ۳-۱، روند انجام فرآیندهای پیش‌پردازش را به ترتیب برای یک نمونه رقم هفت نشان می‌دهد.



شکل ۳-۱: نتیجه‌ی اسکن شده‌ی یک نمونه‌ی تایپی عدد هفت، چرخش ۴۵ درجه‌ی آن خلاف عقربه‌های ساعت و تشکیل مستطیل محاطی آن به منظور تشکیل بردار مشخصه.

۳-۲-۳- استخراج مشخصه‌ها

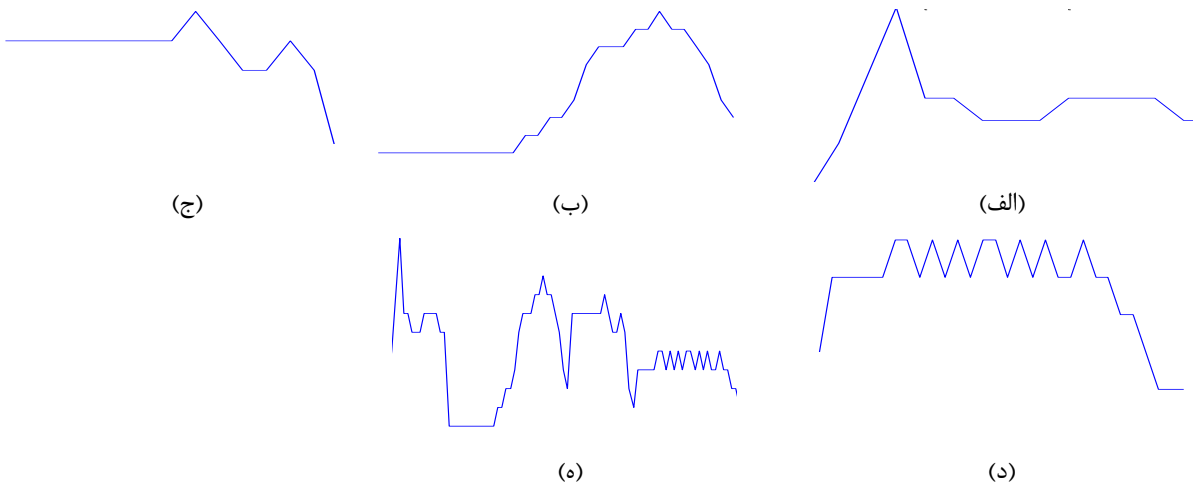
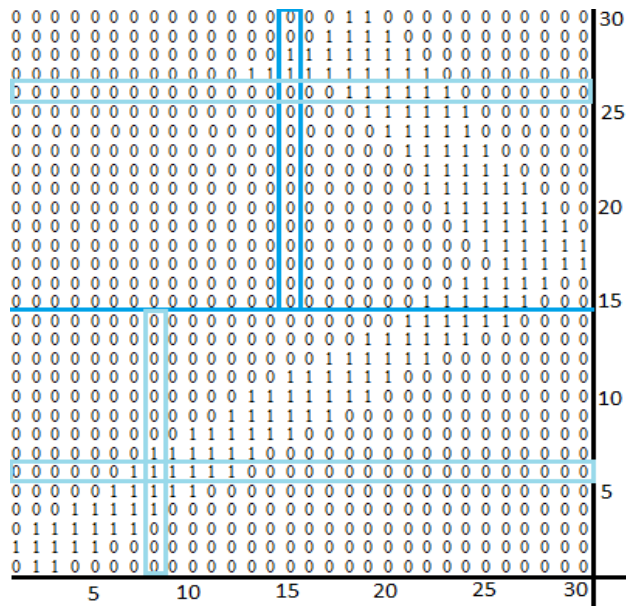
عملکرد سیستم تشخیص ارائه شده در این تحقیق تا حد زیادی وابسته به این مرحله است. بنابراین، مشخصه‌ی استخراج شده باید به گونه‌ای باشد که توصیف‌گر ارقام هر کلاس بوده و بتواند ویژگی منحصر به فردی برای هر دسته از ارقام ارائه دهد. به عبارت دیگر، اندازه و فونت‌های مختلف هر یک از ارقام دهگانه باید دارای بردار مشخصه‌ی مشابهی باشند تا عمل دسته‌بندی با دقت مناسب انجام گیرد.

¹Quadratic

²Cubic

همانطور که در بخش ۳-۲-۱ توضیح داده شد، پیکسل‌های مربوط به نمایش کاراکتر در یک تصویر با مقدار یک و پیکسل‌های پس‌زمینه با مقدار صفر نشان داده می‌شوند. در روش پیشنهادی، به ازای هر کاراکتر، تصویر را از مرکز آن به دو قسمت تقسیم می‌کنیم. سپس، برای استخراج ویژگی‌های ارقام، مجموع تعداد یک‌ها در هریک از سطرها و ستون‌ها در هرکدام از دو نیمه‌ی تصویر محاسبه می‌شود. با توجه به آنکه تصویر هر رقم در مرحله‌ی پیش‌پردازش در اندازه 30×30 ذخیره می‌شود، هر رقم به یک سری زمانی با طول ۹۰ تبدیل می‌گردد (شکل ۳-۲).

در بخش ۳-۲-۲، بیان شد که تصاویر به اندازه‌ی ۴۵ درجه در جهت خلاف حرکت عقربه‌های ساعت، چرخانده می‌شوند. این چرخش، سبب زیاد شدن اختلاف بین کاراکترهای مشابه می‌شود؛ زیرا پس از تقسیم تصویر به دو قسمت از نقطه‌ی میانی آن، پیکسل‌های متفاوتی از کاراکترهای مختلف در این دو نیمه قرار می‌گیرند و این امر، سبب می‌شود ارقام و حروف نسبتاً شبیه به یکدیگر مانند ۸، ۶ و ۹، ۰ یا B، D تفاوت بیشتری در تعداد پیکسل‌های یک آنها در دو نیمه ایجاد شود و عمل بازشناسی آنها راحت‌تر گردد. به عنوان مثال، اگر دو رقم ۰ و ۹ را در نظر بگیریم، در صورت عدم چرخش، با تقسیم تصویر به دو نیمه، نمایش دو رقم در نیمه‌ی بالایی تصویر، مشابه و تعداد یک‌ها تقریباً برابر خواهند بود؛ بنابراین، ممکن است با برابر بودن تعداد یک‌ها در نیمه‌ی پایینی این دو رقم در یکی از فونت‌ها، در هنگام بازشناسی، این دو رقم با یکدیگر اشتباه گرفته شوند؛ اما، اگر چرخش به اندازه‌ی ۴۵ درجه در جهت خلاف عقربه‌های ساعت داشته باشیم، این مشکل به وجود نخواهد آمد. شکل ۳-۳، یک نمونه از هریک از ارقام ۰ و ۹ قبل و بعد از چرخش نشان داده شده‌اند. قبل از چرخش، تعداد پیکسل‌های یک نیمه‌ی بالایی دو رقم ۰ و ۹ تقریباً به تعداد برابر هستند و این امر شباهت بین این دو رقم را زیاد می‌کند. اما رقم ۰ بعد از چرخش، تعداد پیکسل یک بیشتری در نیمه‌ی بالایی تصویر دارد و سبب می‌شود شناسایی این ارقام، با دقت بالاتری صورت گیرد؛ به طوری که این چرخش ۴۵ درجه، سبب افزایش ۲۰ درصدی در دقت بازشناسی در داده‌های آزمایش شده است. شکل ۳-۴، بردارهای ویژگی حاصل از این دو رقم را قبل از چرخش و بعد از آن نشان می‌دهد. همانطور که مشاهده می‌شود، شباهت این دو بردار ویژگی قبل از چرخش، بیشتر از شباهت آنها بعد از چرخش است و دو سری زمانی، قبل از چرخش نقاط مشترک بیشتری نسبت به بعد از چرخش دارند که این امر تشخیص صحیح این دو رقم در فونت‌های مختلف را با مشکل مواجه می‌کند.



۲	۴	۷	۱۰	۶	...	۰	...	۱	۲	...	۶	...	۶	۶	۵	...	۱	۳	۳	...	۴	۳	۴	...
---	---	---	----	---	-----	---	-----	---	---	-----	---	-----	---	---	---	-----	---	---	---	-----	---	---	---	-----

(ی)

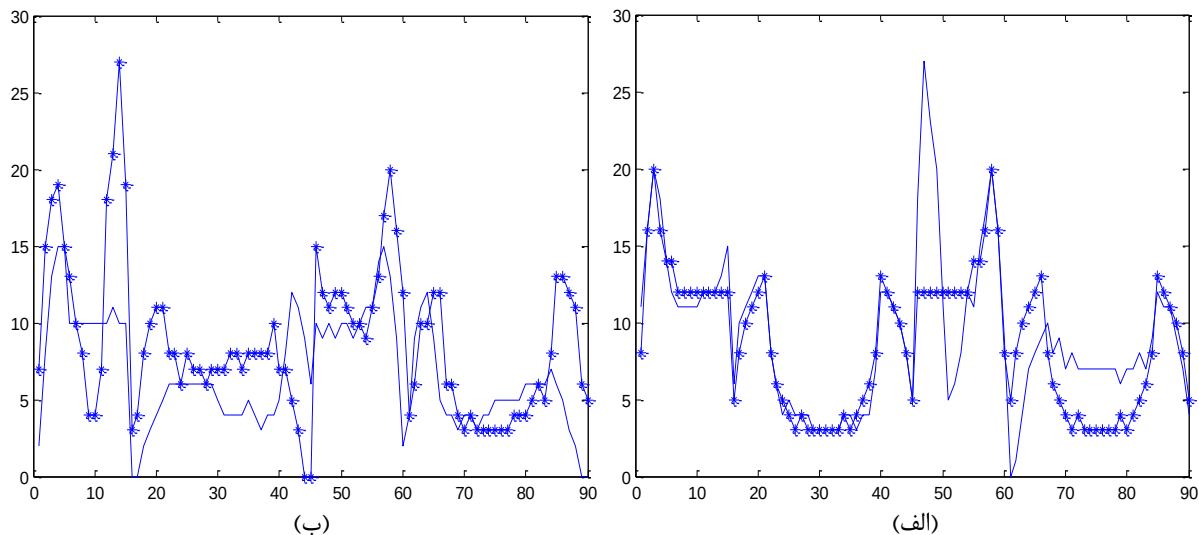
شکل ۳-۲: نمونه‌ای از رقم هفت با فونت Times New Roman: (الف) سری زمانی حاصل از مجموع یک‌ها در سطرهای متوالی در نیمه‌ی بالایی تصویر؛ (ب) سری زمانی حاصل از مجموع یک‌ها در ستون‌های متوالی در نیمه‌ی بالایی تصویر؛ (ج) سری زمانی حاصل از مجموع یک‌ها در سطرهای متوالی در نیمه‌ی پایینی تصویر؛ (د) سری زمانی حاصل از مجموع یک‌ها در ستون‌های متوالی در نیمه‌ی پایینی تصویر؛ (ه) سری زمانی حاصل از ترکیب سری‌های زمانی الف، ب، ج و د؛ (ی) بردار مشخصه استخراج شده. (اعدادی که با رنگ خاکستری در این بردار مشخص شده‌اند، حاصل جمع مقادیر هریک از سطرها و ستون‌های مشخص شده در تصویر رقم ۷ هستند).



(ب)

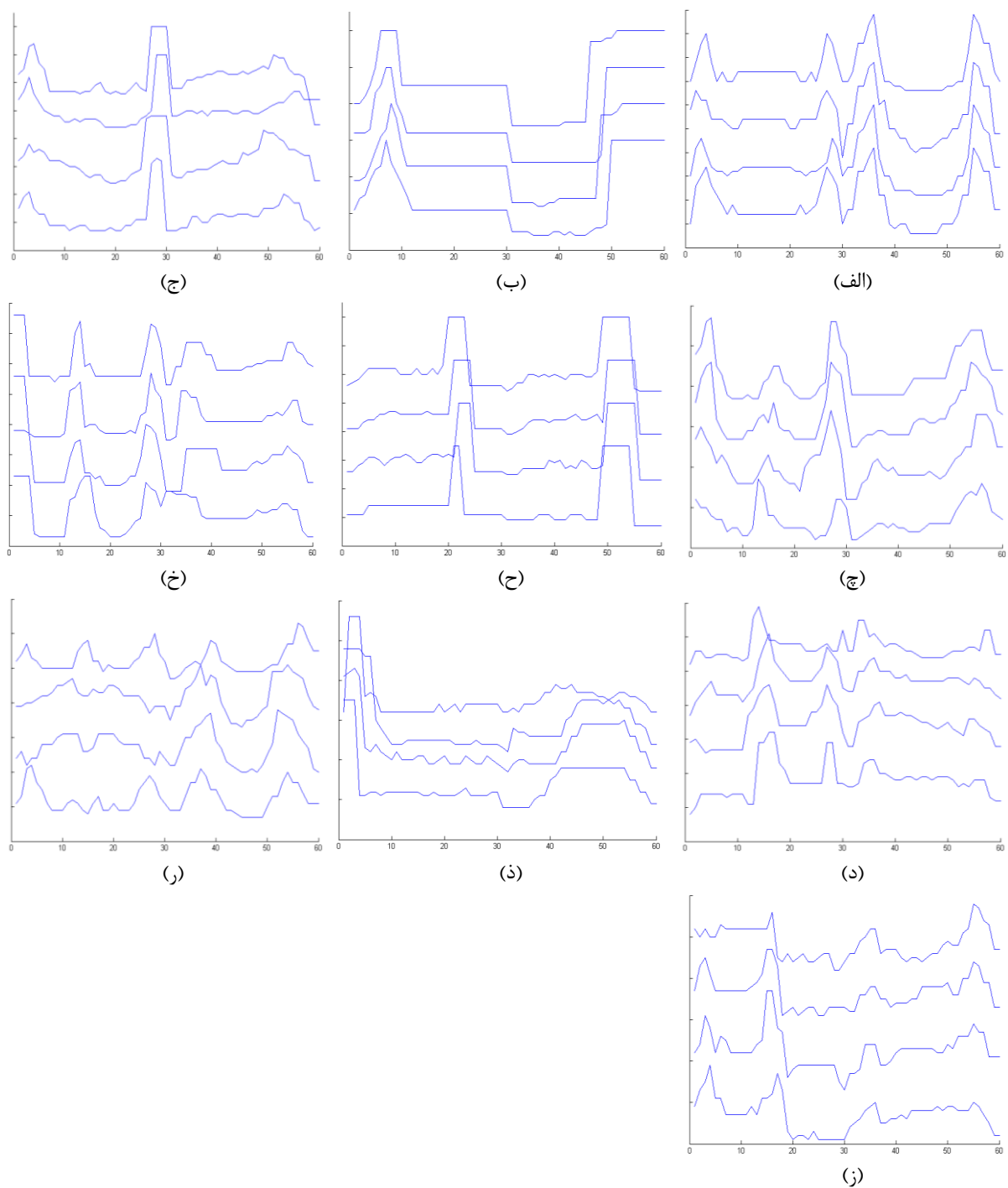
(الف)

شکل ۳-۳: شباهت نیمه‌ی بالایی دو رقم 0 و 9، الف) قبل و ب) بعد از چرخش ۴۵ درجه در جهت خلاف عقربه‌های ساعت.

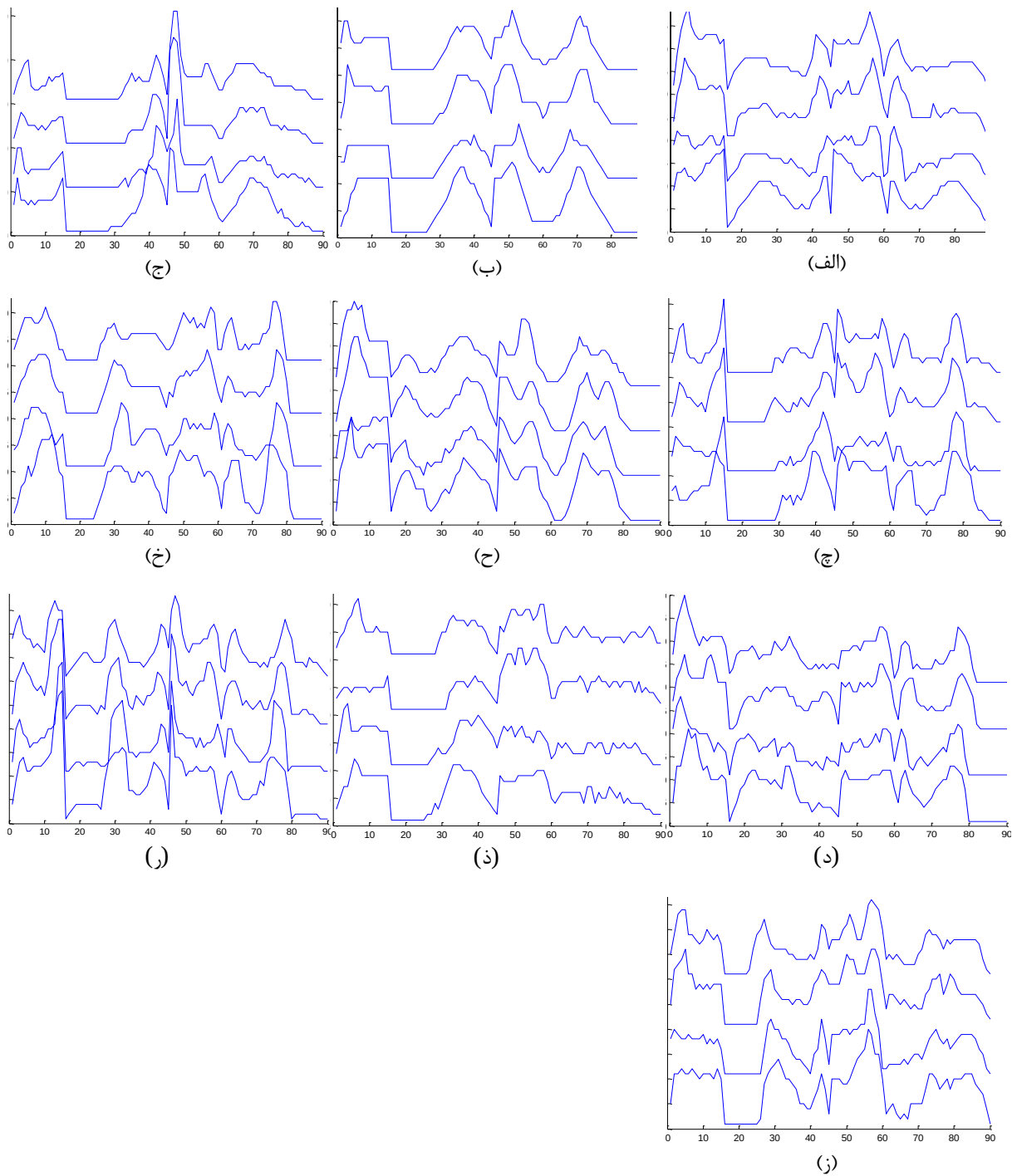


شکل ۳-۴: الف) سری زمانی حاصل از بردار ویژگی ارقام ۰ و ۹ قبل از چرخش، ب) بعد از چرخش.

در شکل ۳-۵، سری‌های زمانی (بردار ویژگی) استخراج شده برای چهار نمونه‌ی (فونت) مختلف از هریک از ارقام صفر تا ۹، قبل از چرخش نمایش داده شده است. شکل ۳-۶، این سری‌های زمانی را بعد از چرخش ۴۵ درجه نشان می‌دهد. همانطور که مشاهده می‌شود سری‌های زمانی کلاس‌های مختلف به خصوص کلاس ارقام مشابه بعد از چرخش، تفاوت بیشتری با یکدیگر دارند.

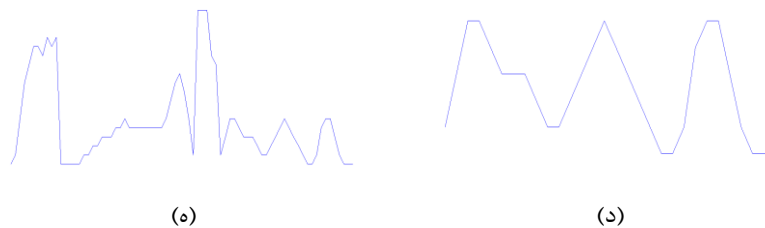
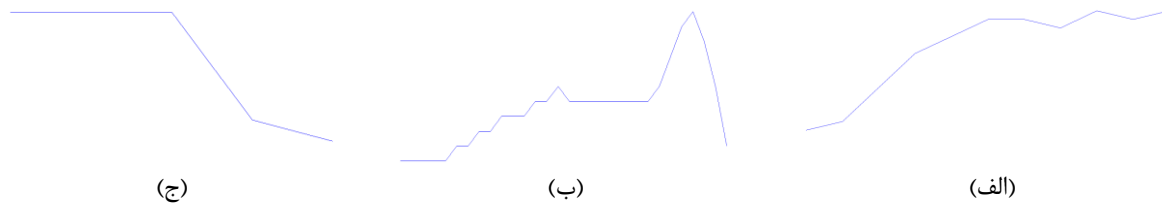
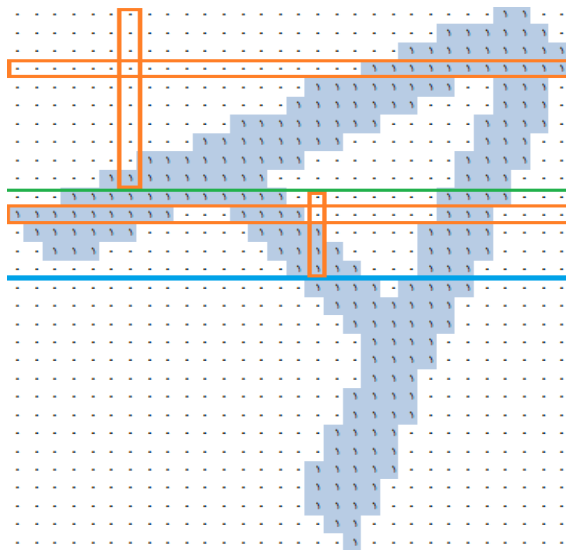


شکل ۳-۵: سری زمانی حاصل از بردار ویژگی استخراج شده برای ارقام الف (۰، ب (۱، ج (۲، چ (۳، ح (۴، خ (۵، د (۶، ذ (۷، ر (۸، ز (۹، قبل از چرخش. در هر یک از این شکل‌ها، سری زمانی مربوط به چهار نمونه مختلف از هر رقم آورده شده است که به ترتیب از بالا به پایین مربوط به فونت‌های Century bold, Arial regular, Century و Gothic regular, Estrangelo Eddesa bold می‌باشند. (محور افقی نمایانگر تعداد ویژگی‌ها است و محور عمودی مقدار هر ویژگی را در هریک از چهار نمودار نشان می‌دهد).



شکل ۳-۶: سری زمانی حاصل از بردار ویژگی استخراج شده برای ارقام الف (۰، ب (۱، ج (۲، چ (۳، ح (۴، خ (۵، د (۶، ذ (۷، ر (۸، ز (۹ پس از چرخش. در هر یک از این شکل‌ها سری زمانی مربوط به چهار نمونه مختلف از هر رقم آورده شده است که به ترتیب از بالا به پایین مربوط به فونت‌های Century bold, Arial regular, Century Gothic regular, و Estrangelo Eddesa bold می‌باشند. (محور افقی نمایانگر تعداد ویژگی‌ها است و محور عمودی مقدار هر ویژگی را در هر یک از چهار نمودار نشان می‌دهد).

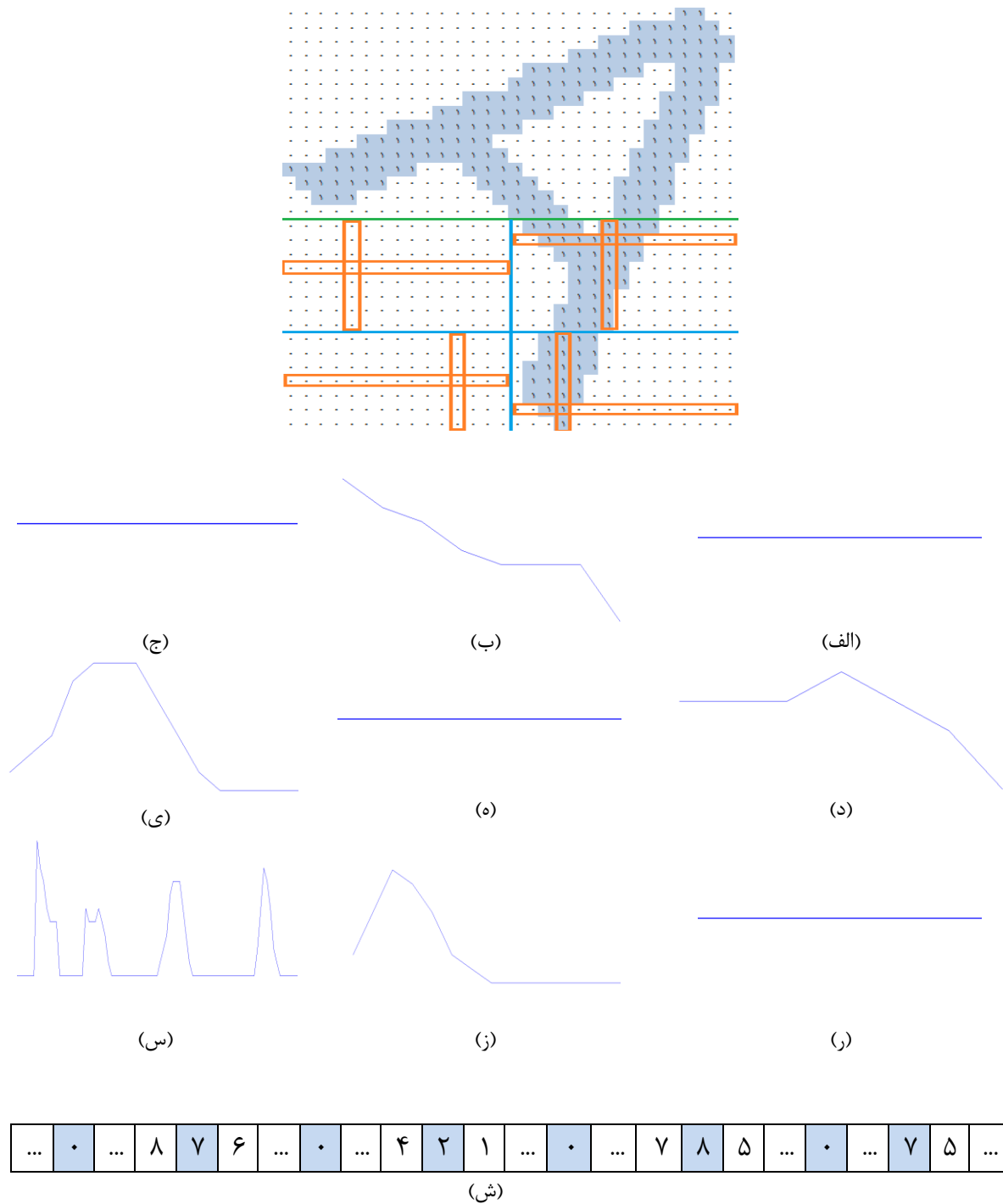
تغییر فونت در یک رقم، سبب تغییرات کمی در شکل نوشتاری ارقام مشابه آن می‌گردد؛ در حالی که تغییرات فونت در یک حرف، تفاوت‌های زیادی در نمایش آن حرف مثلاً با افزوده شدن سریف‌ها به حروف ایجاد می‌کند؛ بنابراین، ویژگی‌های جمع سطر و ستون، برای متمایز کردن حروف کافی نیستند. برای استخراج ویژگی‌های حروف، دو مجموعه ویژگی در نظر گرفته‌ایم. مجموعه‌ی اول ویژگی‌ها بدین ترتیب به دست می‌آیند: با تقسیم نیمه‌ی بالایی تصویر به نسبت $\frac{1}{3}$ و $\frac{2}{3}$ ، جمع مقادیر سطرها و ستون‌های این دو ناحیه را در این نیمه حساب می‌کنیم. از آنجایی که تصاویر در اندازه‌ی 30×30 هستند، ۷۵ ویژگی از بردار ویژگی حاصل می‌شود. برای یافتن مجموعه‌ی دوم ویژگی‌ها، نیمه‌ی پایینی تصاویر را در دو جهت افقی و عمودی به دو قسمت تقسیم می‌کنیم. بنابراین، چهار ناحیه‌ی مساوی خواهیم داشت که جمع سطرها و ستون‌های هریک از این چهار ناحیه را محاسبه می‌کنیم. در این صورت، ۹۰ ویژگی به دست می‌آید. با ترکیب این دو مجموعه ویژگی‌ها، یک بردار ویژگی (سری زمانی) ۱۶۵ تایی خواهیم داشت. شکل‌های ۳-۷ و ۳-۸، به ترتیب مجموعه‌ی اول و دوم ویژگی‌ها و سری زمانی مربوط به یک حرف A را نمایش می‌دهند. شکل ۳-۹، سری زمانی ۱۶۵ تایی نهایی حرف A و بردار مشخصه متناظر با آن را شامل می‌شود. شکل ۳-۱۰، سری‌های زمانی حاصل از ۱۶۵ ویژگی را برای چهار نمونه از فونت‌های مختلف هر یک از حروف بزرگ A تا F پیش از چرخش ۴۵ درجه نشان می‌دهد. همانطور که مشاهده می‌شود این سری‌های زمانی شباهت زیادی به هم داشته و برای حروف مختلف متمایز نیستند؛ به خصوص حروف با شکل نوشتاری مشابه مانند B و D که سری‌های زمانی آنها شباهت بیشتری با یکدیگر دارند. شکل‌های ۳-۱۱ و ۳-۱۲، نمایش دهنده‌ی سری‌های زمانی حاصل ۱۶۵ ویژگی مربوط به هریک از حروف بزرگ و کوچک برای چهار نوع فونت مختلف پس از چرخش ۴۵ درجه هستند.



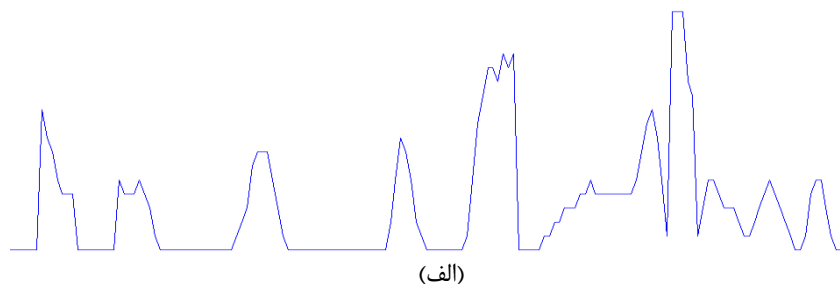
۲	۶	۹	۱۱	۱۱	...	۰	۱	۱	۲	...	۱۶	۱۶	۱۴	...	۴	۴	۳	۲	...
---	---	---	----	----	-----	---	---	---	---	-----	----	----	----	-----	---	---	---	---	-----

(ی)

شکل ۳-۷: نمونه‌ای از کاراکتر A با فونت Arial: الف) سری زمانی حاصل از مجموع یک‌ها در سطرهاى متوالی ۲/۳ نیمه‌ی بالایی تصویر؛ ب) سری زمانی حاصل از مجموع یک‌ها در ستون‌های متوالی ۲/۳ نیمه‌ی بالایی تصویر؛ ج) سری زمانی حاصل از مجموع یک‌ها در سطرهاى متوالی ۱/۳ نیمه‌ی بالایی تصویر؛ د) سری زمانی حاصل از مجموع یک‌ها در ستون‌های متوالی ۱/۳ نیمه‌ی بالایی تصویر؛ ه) سری زمانی حاصل از ترکیب سری‌های الف و ب و ج و د؛ ی) بردار مشخصه‌ی استخراج شده. (اعدادی که با رنگ خاکستری در این بردار مشخص شده‌اند، حاصل جمع مقادیر هریک از سطرها و ستون‌های مشخص شده در تصویر حرف A هستند).



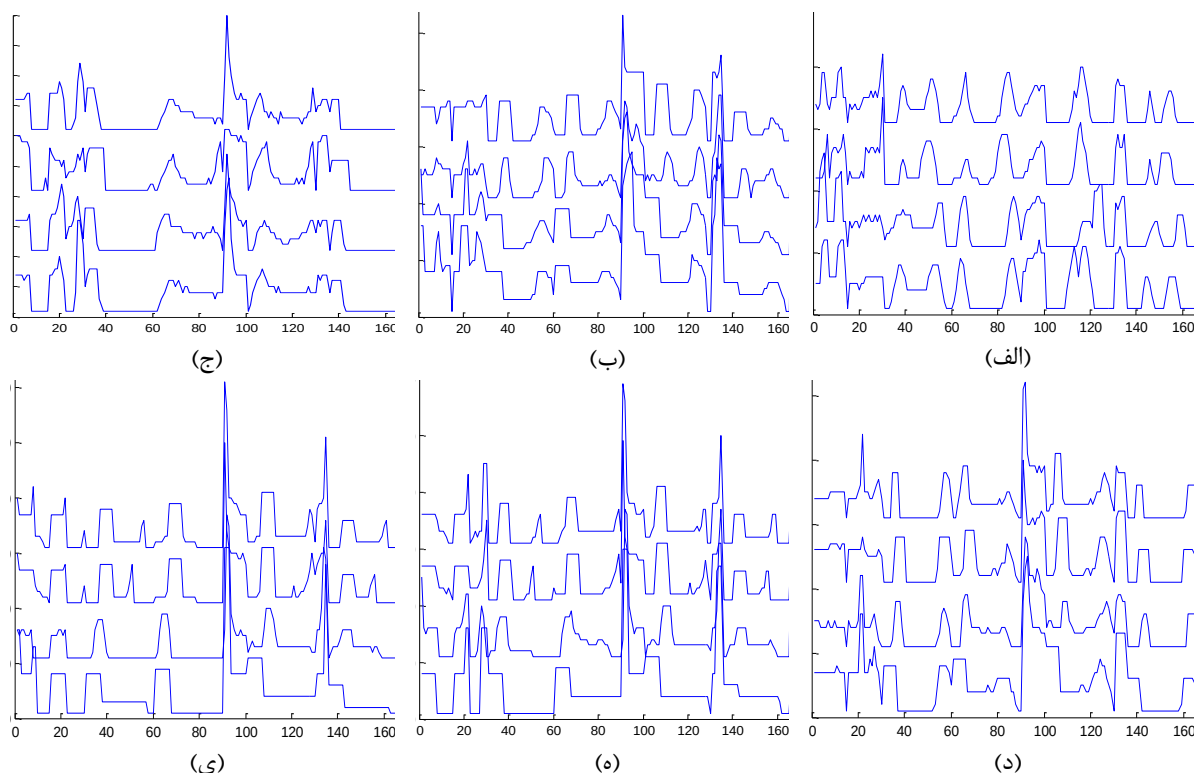
شکل ۳-۸: نمونه‌ای از کاراکتر A با فونت Arial: الف، ب، ج، د) سری زمانی حاصل از مجموع یک‌ها در سطرها، متوالی ناحیه‌ی اول، دوم، سوم و چهارم نیمه‌ی پایینی تصویر؛ ه، ی، ر، ز) سری زمانی حاصل از مجموع یک‌ها در ستون‌های متوالی ناحیه‌ی اول، دوم، سوم و چهارم نیمه‌ی پایینی تصویر؛ س) سری زمانی حاصل از ترکیب سری‌های زمانی الف، ب، ج، د، ه، ی، ر و ز؛ ش) بردار مشخصه‌ی استخراج شده. (اعدادی که با رنگ خاکستری مشخص شده مجموع مقادیر سطرها و ستون‌های مشخص شده در تصویر حرف A هستند.)



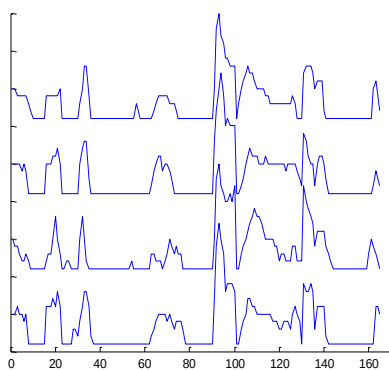
...	۰	...	۸	۷	۶	...	۰	...	۴	۲	۱	...	۲	۶	۹	۱۱	...	۵	۴	۱	۰	۰	۰
-----	---	-----	---	---	---	-----	---	-----	---	---	---	-----	---	---	---	----	-----	---	---	---	---	---	---

(ب)

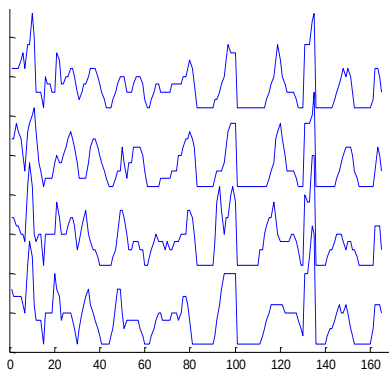
شکل ۳-۹: الف) سری زمانی ۱۶۵ تایی حاصل از بردار ویژگی نمونه کاراکتر A نشان داده شده در شکل‌های ۳-۷ و ۳-۸. ب) بردار ویژگی نمونه کاراکتر A مذکور.



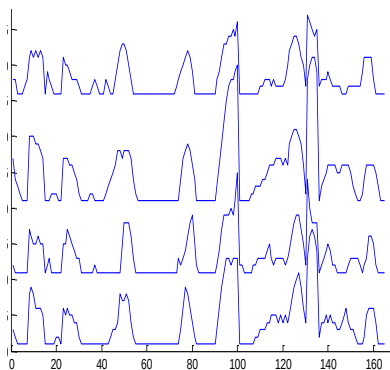
شکل ۳-۱۰: سری زمانی‌های (الف) تا (د) به ترتیب سری زمانی حاصل از بردار ویژگی استخراج شده برای حروف A تا F را قبل از چرخش ۴۵ درجه نمایش می‌دهند. در هر یک از این شکل‌ها سری زمانی مربوط به چهار نمونه‌ی مختلف از هر حرف آورده شده است که به ترتیب از پایین به بالا مربوط به فونت‌های 'Lucida Sans Typewriter'، 'Bodoni MT'، 'Rockwell' و 'Palatino' می‌باشند. (محور افقی نمایانگر تعداد ویژگی‌ها است و محور عمودی مقدار هر ویژگی را در هر یک از چهار نمودار نشان می‌دهد).



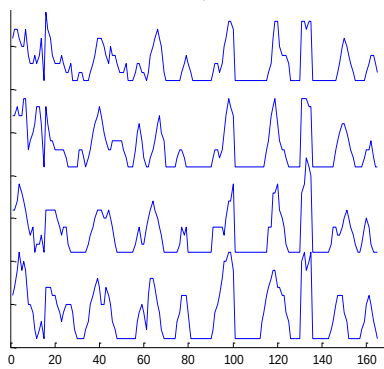
(پ)



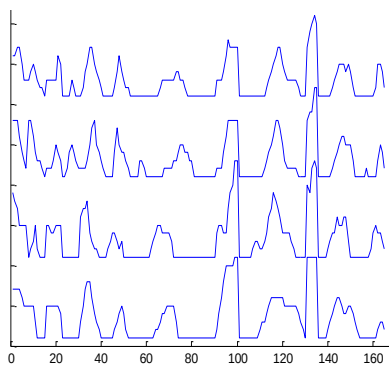
(ب)



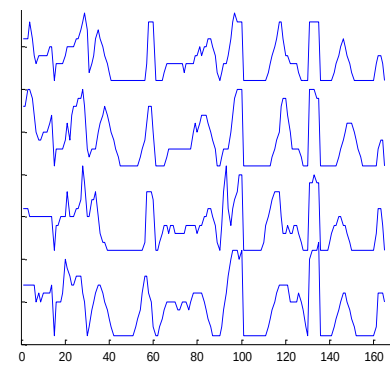
(الف)



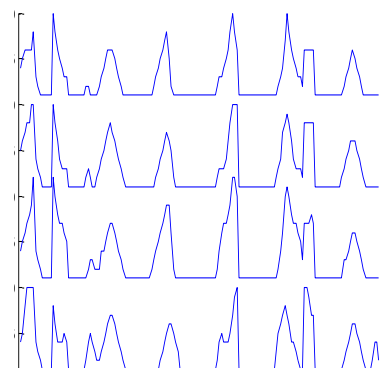
(چ)



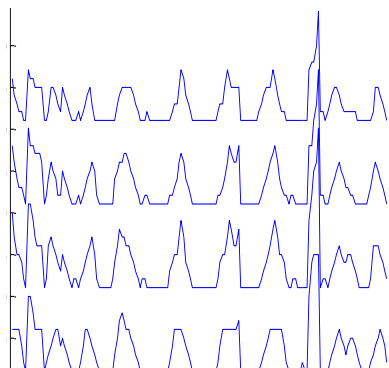
(ث)



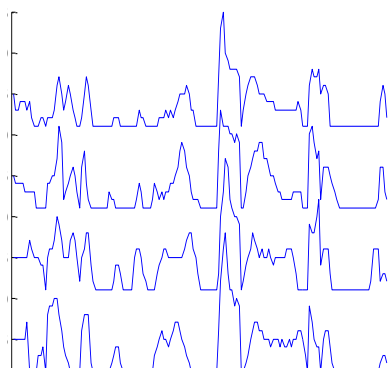
(ت)



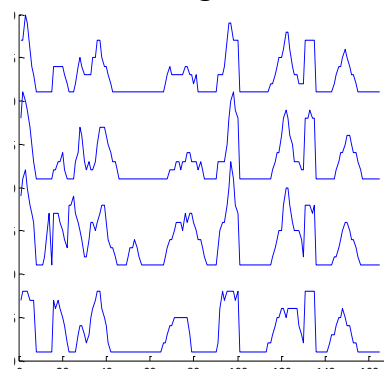
(خ)



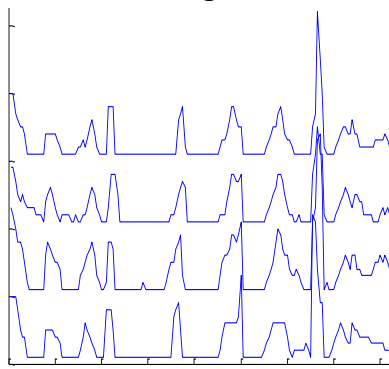
(ح)



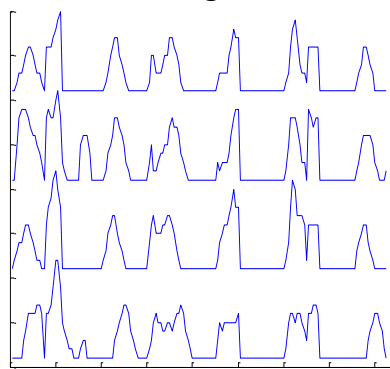
(ج)



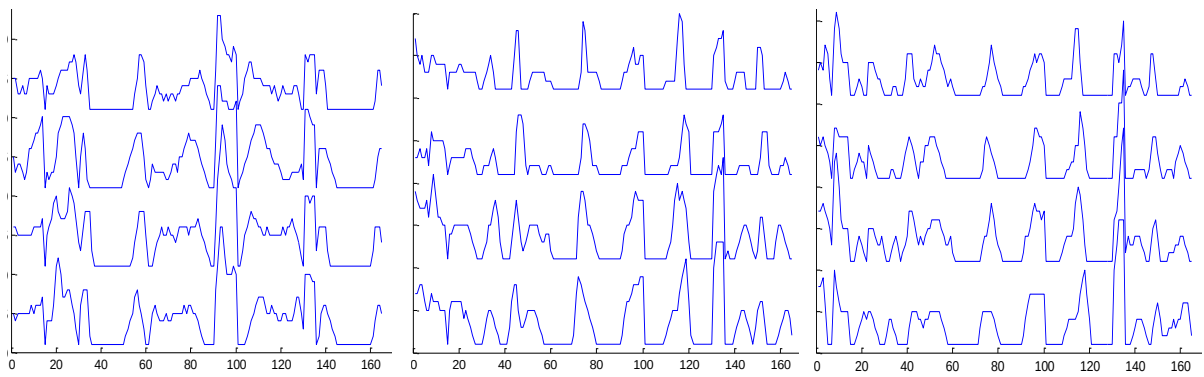
(ر)



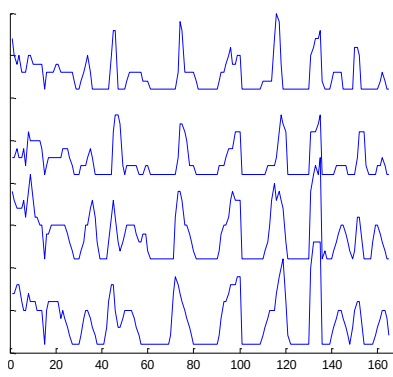
(ذ)



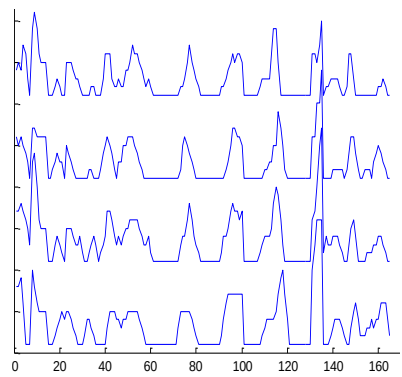
(د)



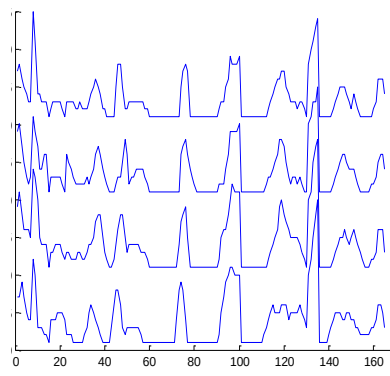
(ش)



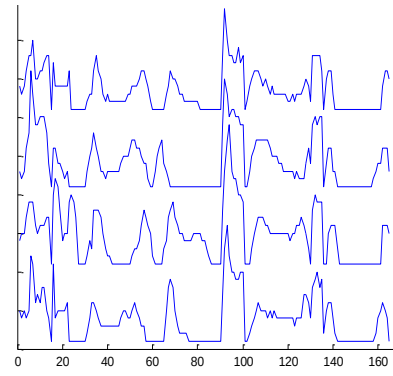
(س)



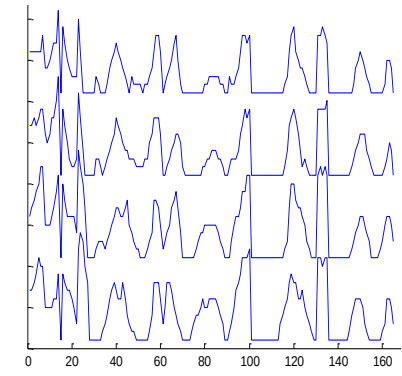
(ز)



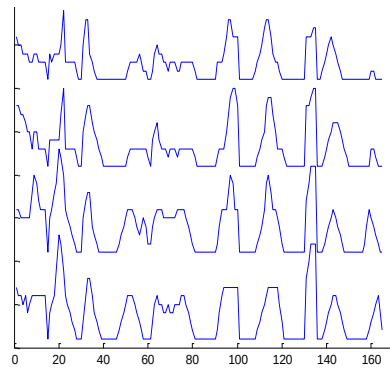
(ط)



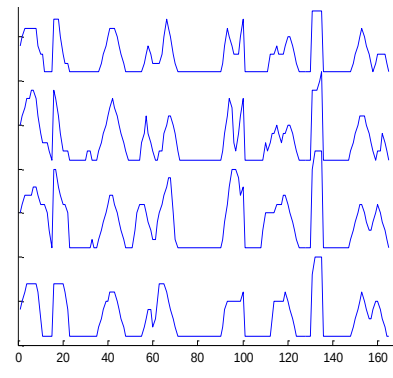
(ض)



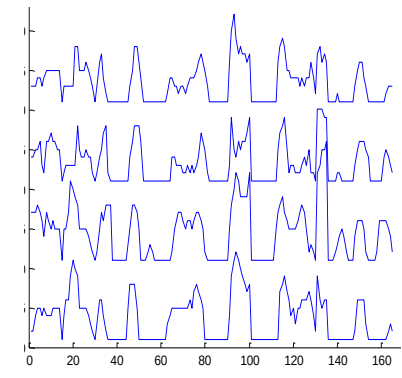
(ص)



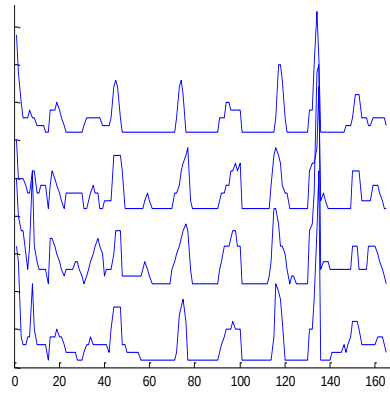
(غ)



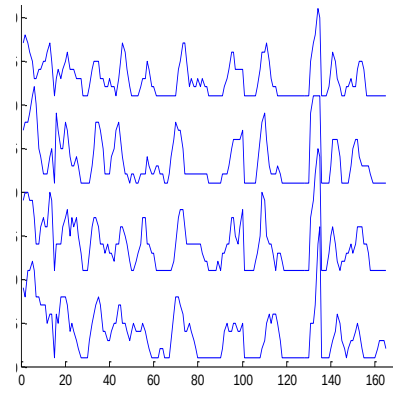
(ع)



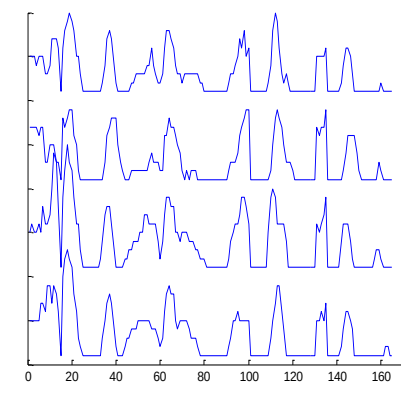
(ظ)



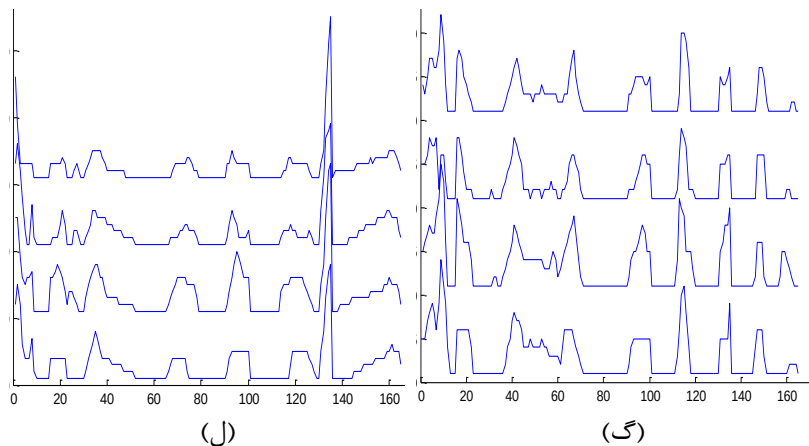
(ک)



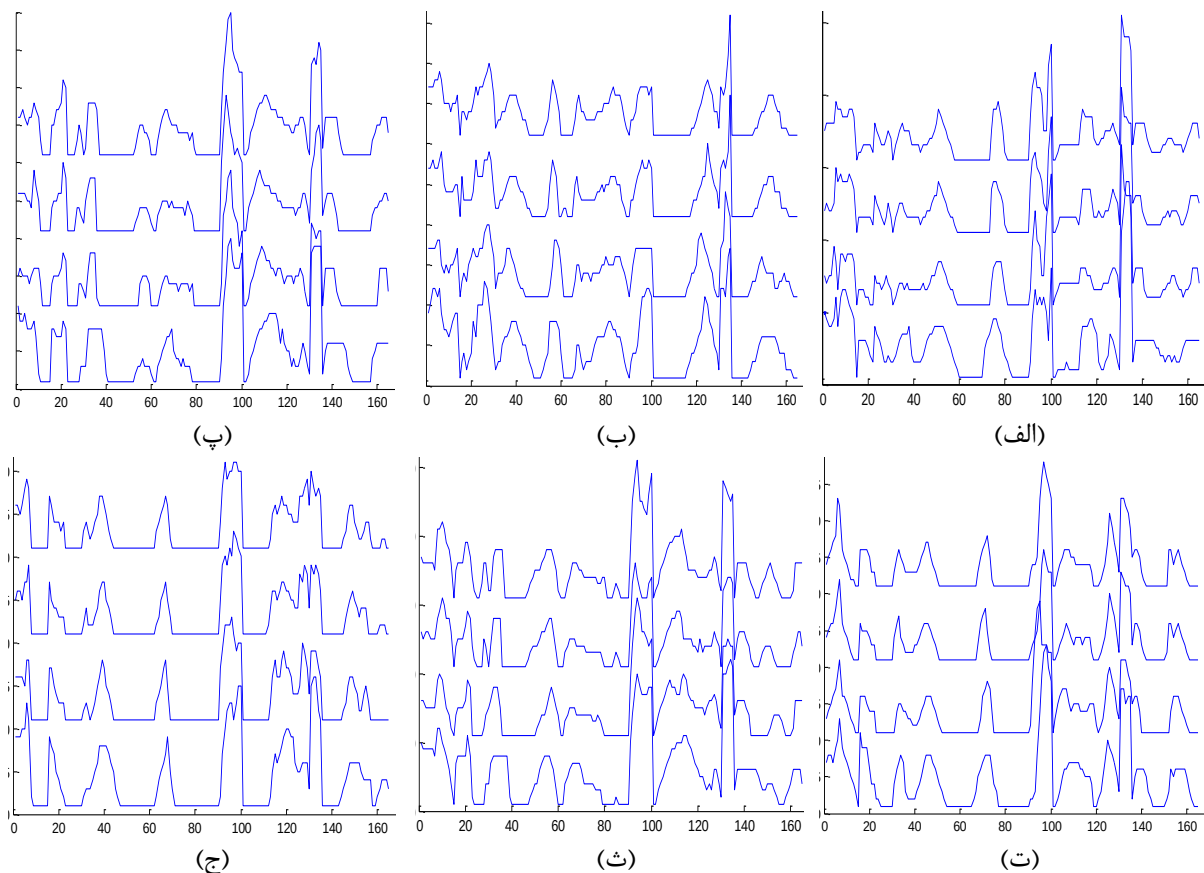
(ق)

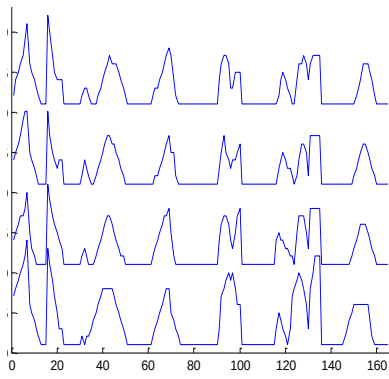


(ف)

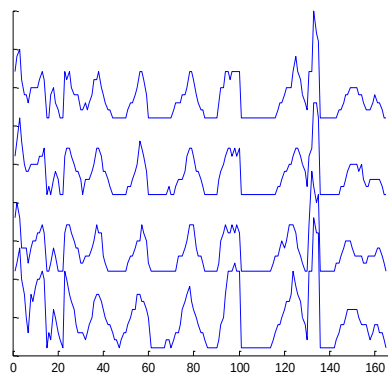


شکل ۳-۱۱: سری زمانی‌های (الف) تا (ل) به ترتیب سری زمانی حاصل از بردار ویژگی استخراج شده برای حروف A تا Z را پس از چرخش نمایش می‌دهند. در هر یک از این شکل‌ها سری زمانی مربوط به چهار نمونه‌ی مختلف از هر حرف آورده شده است که به ترتیب از پایین به بالا مربوط به فونت‌های Palatino، Bodoni MT، Rockwell و Lucida Sans Typewriter می‌باشند. (محور افقی نمایانگر تعداد ویژگی‌ها است و محور عمودی مقدار هر ویژگی را در هریک از چهار نمودار نشان می‌دهد.)

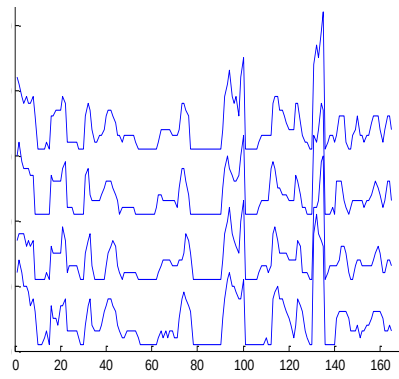




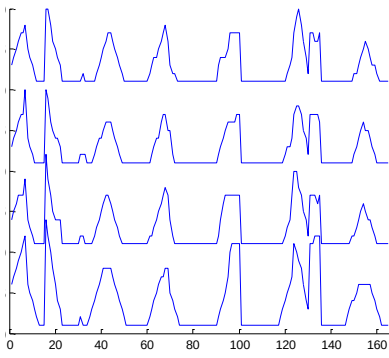
(خ)



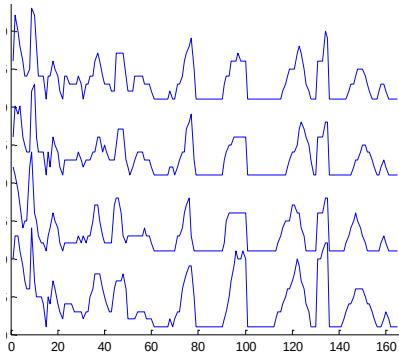
(ح)



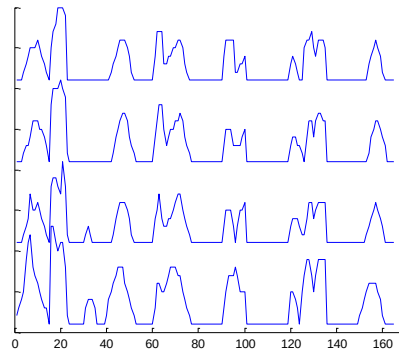
(ج)



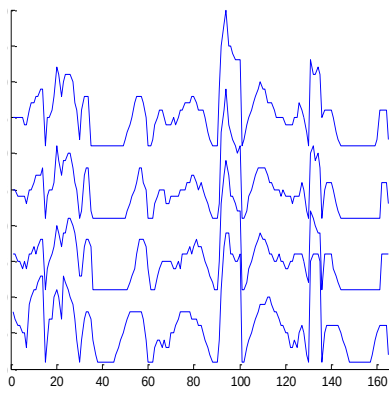
(ر)



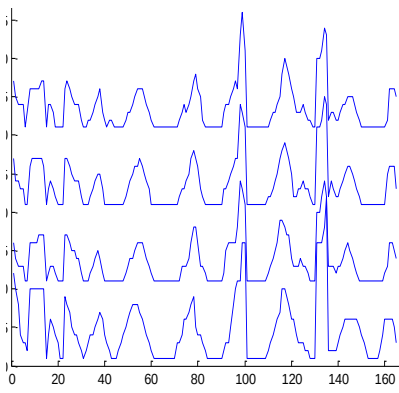
(ذ)



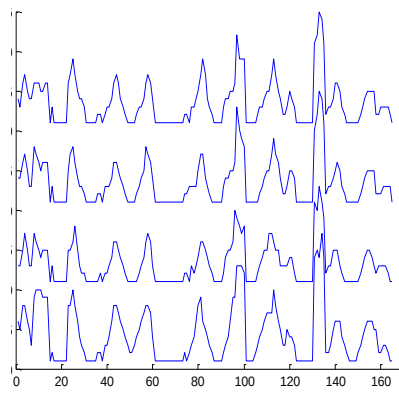
(د)



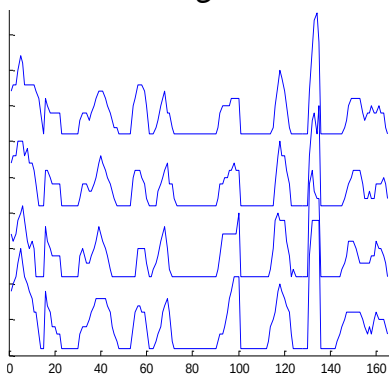
(ش)



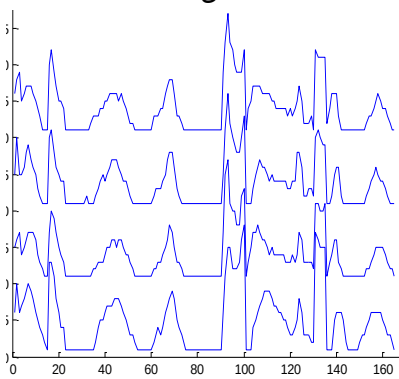
(س)



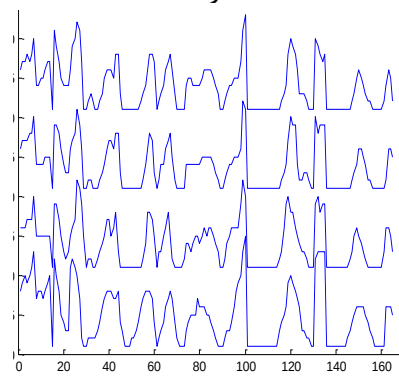
(ز)



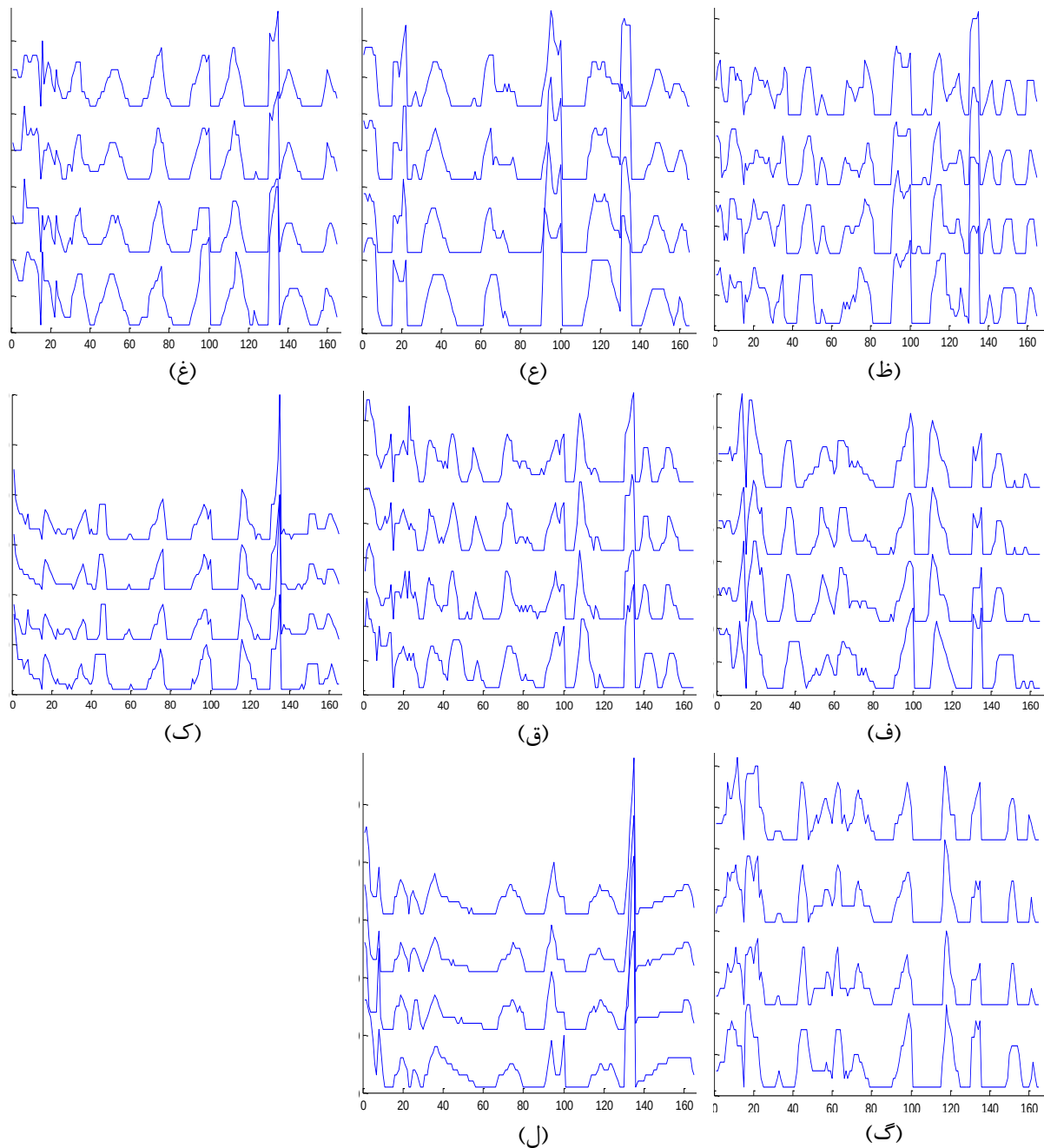
(ط)



(ض)



(ص)

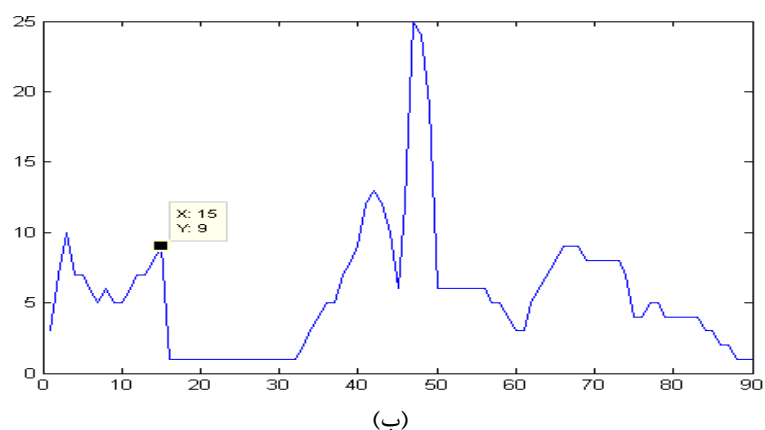
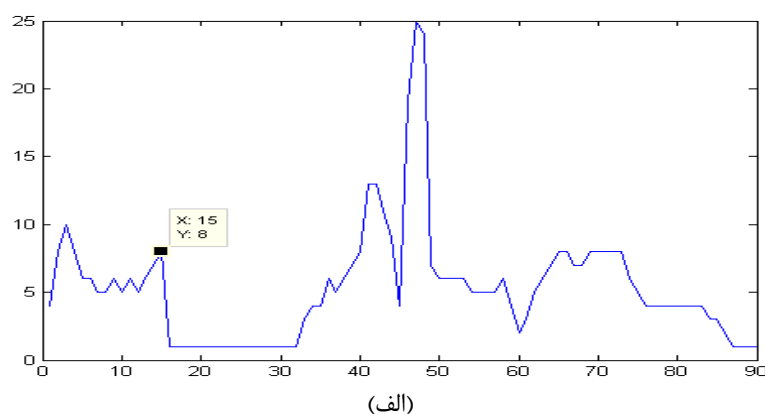


شکل ۳-۱۲: سری زمانی‌های (الف) تا (ل) به ترتیب سری زمانی حاصل از بردار ویژگی استخراج شده برای حروف a تا z را پس از چرخش نمایش می‌دهند. در هر یک از این شکل‌ها سری زمانی مربوط به چهار نمونه‌ی مختلف از هر حرف آورده شده است که به ترتیب از پایین به بالا مربوط به فونت‌های Palatino، Bodoni MT، Rockwell و Lucida Sans Typewriter می‌باشند. (محور افقی نمایانگر تعداد ویژگی‌ها است و محور عمودی مقدار هر ویژگی را در هریک از چهار نمودار نشان می‌دهد).

۳-۲-۴- طبقه‌بندی و شناسایی

۳-۲-۴-۱- معیار شباهت

سری‌های زمانی حاصل از دو رقم ۲ با فونت Times new roman در اندازه ۲۶ و ۲۸ در شکل ۳-۱۳، نشان داده شده‌اند. دو نمودار به یکدیگر شبیه هستند با وجودی که اولین نقطه مینیمم سری زمانی شکل (ب) نسبت به (الف) مقدار بیشتری دارد و این نمودار دارای شکستگی‌های بیشتری نسبت به نمودار (الف) است.



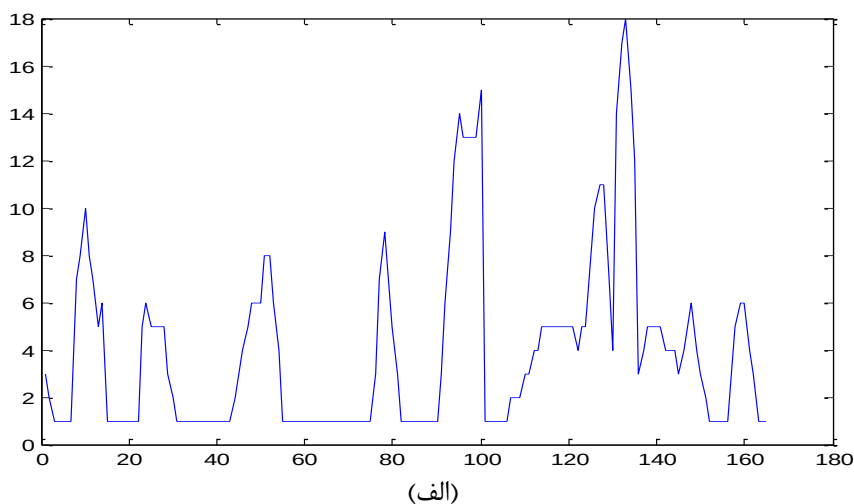
شکل ۳-۱۳: سری زمانی حاصل از دو نمونه از رقم ۲ با فونت Times new roman، (الف) در اندازه ۲۶، (ب) در اندازه ۲۸.

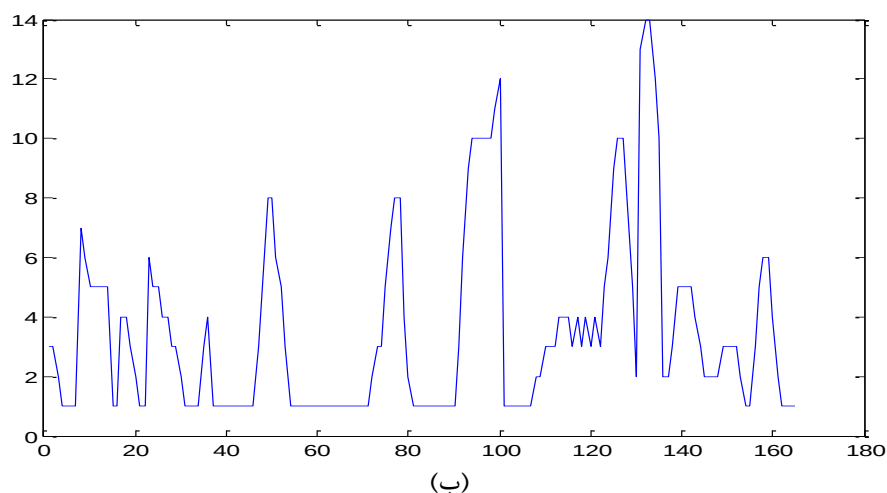
حال اگر بخواهیم این دو سری زمانی را با فاصله‌ی اقلیدسی با یکدیگر مقایسه کنیم، از آنجایی که نمودارها جابجا شده‌اند (مثلاً در $x=15$ مقدار y ها با یکدیگر متفاوت هستند)، فاصله اقلیدسی عدد ۳۲ (مقداری غیر صفر) را به عنوان فاصله‌ی بین این دو نمودار بر خواهد گرداند؛ در نتیجه دو نمودار مشابه

نخواهند بود و قادر نیست ارقام یک دسته را با وجود شباهت بین آنها مشابه تشخیص دهد. بنابراین نیاز به معیار شباهتی متناسب با کاربرد داریم که به جای محاسبه‌ی فاصله، شباهت بین نمونه‌ها را بررسی کند.

شکل ۳-۱۴، سری‌های زمانی دو نوع فونت مختلف از حرف A در اندازه‌ی ۲۴ را نشان می‌دهند. همانطور که مشاهده می‌شود این دو سری زمانی شبیه به هم هستند اما علاوه بر متفاوت بودن نقاط ماکسیمم و مینیمم، تفاوت‌هایی دیگر مانند شکستگی بین نقاط ماکسیمم و مینیمم نیز بین آنها دیده می‌شود. به دلیل وجود این تفاوت‌ها، معیار فاصله‌ای مانند فاصله‌ی اقلیدسی قادر نخواهد بود شناسایی را با دقت بالایی انجام دهد؛ چرا که اندازه‌گیری فاصله‌ی بین این دو نمودار، رقم بزرگی را به عنوان میزان فاصله‌ی این دو نمودار برخواهد گرداند اما با اعمال یک معیار شباهت متناسب با کاربرد، رقم برگردانده شده مقدار کوچکی خواهد داشت که نشان از شباهت این دو نمودار است.

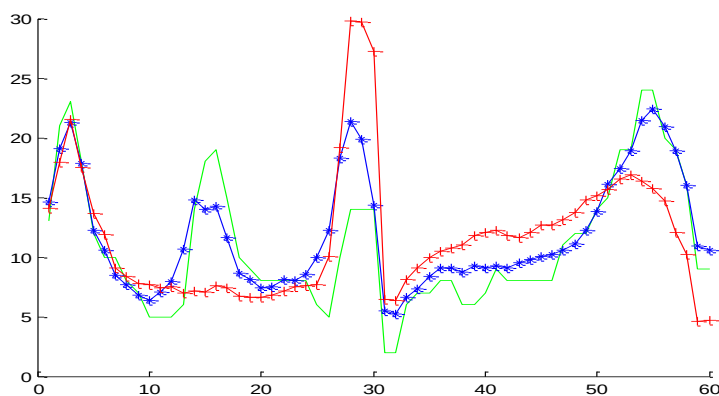
با بیان مثال دیگری، عدم موفقیت فاصله‌ی اقلیدسی را در تشخیص ارقام شرح می‌دهیم. فرض کنید رقم سه با فونت Arial و سایز ۲۸ را می‌خواهیم در دسته‌ی مناسب مربوط به آن قرار دهیم. این رقم باید با وزن برنده‌ی تمام ارقام ۰ تا ۹ مقایسه گردد (با به دست آوردن فاصله‌ی اقلیدسی) و این فاصله با هر کدام از وزن‌ها کمتر بود، این رقم به آن دسته تعلق دارد. در شکل ۳-۱۵، فاصله‌ی سری زمانی رقم ۳ با وزن برنده کلاس رقم ۳، ۵۲٫۸۹ و با وزن برنده رقم ۲، ۴۴٫۴۸ است؛ بنابراین، فاصله رقم ۳ با کلاس رقم ۲ کمتر بوده و اشتباهاً در این کلاس قرار می‌گیرد. درحالی که همانطور که مشاهده می‌گردد دو نمودار





شکل ۳-۱۴: بردار ویژگی نمونه‌ای از حرف A در اندازه ۲۴. الف) فونت Arial، ب) فونت Cambria.

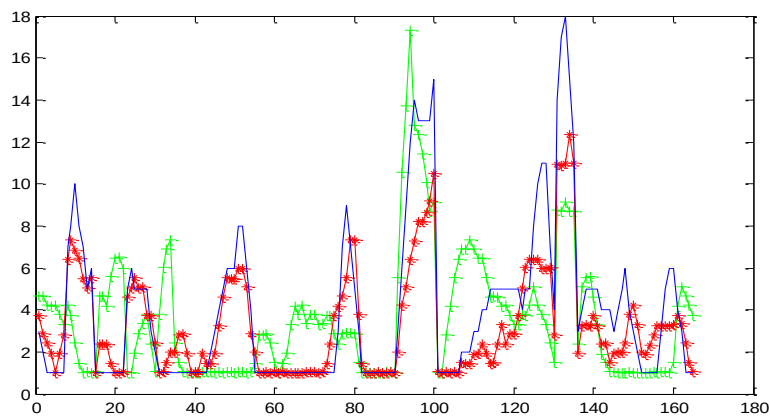
اصلی و نمودار با علامت * به یکدیگر شباهت بیشتری داشته و از دید ناظر انسانی فاصله‌ی کمتری با یکدیگر دارند؛ اما فاصله اقلیدسی، فاصله‌ی نمودار اصلی با نمودار با علامت + را کمتر محاسبه کرده است. این کلاس بندی نادرست، بر مناسب نبودن فاصله اقلیدسی برای مقایسه‌ی ارقام تاکید می‌کند. بنابراین در این مثال نیز به معیار شباهت مناسب نیازمندیم.



شکل ۳-۱۵: نمودار اصلی سری زمانی رقم ۳، نمودار با علامت * وزن برنده‌ی کلاس رقم ۳ و نمودار با علامت + وزن برنده رقم ۲ است.

در شکل ۳-۱۶، نمودار سری زمانی حرف A با فونت Arial به همراه نمودار برنده‌ی کلاس A (نمودار *) و برنده‌ی کلاس C (نمودار +) نشان داده شده است. فاصله‌ی اقلیدسی، فاصله‌ی بین این نمودار را با کلاس A، ۱۱۳،۷۸ و با کلاس C مقدار ۶۷،۸۴ اندازه‌گیری کرد. این درحالی است که شباهت یا

میزان نزدیکی نمودار حرف A به کلاس A بیشتر است تا کلاس C. (به نمودارها دقت کنید.) این مقادیر بر این نکته تاکید می کنند که به جای فاصله ی اقلیدسی به یک معیار شباهت برای دسته بندی نیازمندیم.



شکل ۳-۱۶: نمودار اصلی سری زمانی حرف A، نمودار با علامت * وزن برنده ی کلاس حرف A و نمودار با علامت + وزن برنده حرف C است.

در شکل های ۳-۶، ۳-۱۱ و ۳-۱۲ نمودار سری های زمانی چهار نوع فونت مختلف برای هریک از ارقام و حروف نشان داده شده اند. هریک از این شکل ها، نشان می دهند که بردارهای ویژگی ارقام و حروف مختلف دارای طول برابر هستند اما ممکن است دامنه ی تغییرات یکسانی نداشته باشند. ضمناً موقعیت مکانی نقاط ماکزیمم یا مینیمم محلی در بردار ویژگی فونت های مختلف یک رقم و یک حرف نیز ممکن است یکسان نباشند. بر این اساس با استفاده از یک معیار فاصله، نظیر فاصله اقلیدسی، نمی توان با دقت مناسب ارقام و حروف مختلف را با یکدیگر مقایسه کرد. اما از آنجایی که بین بردار ویژگی ارقام و حروف یک کلاس، با توجه به شکل های ۳-۶، ۳-۱۱ و ۳-۱۲، شباهت هایی وجود دارد، لازم است با استفاده از یک معیار شباهت مناسب به مقایسه و دسته بندی ارقام و حروف پرداخت.

با بیان مثال های مذکور، می توان گفت که فاصله ی اقلیدسی در تشخیص ارقام و حروف با اندازه و فونت های مختلف موفق عمل نمی کند و به یک معیار شباهت نیاز داریم. در این پایان نامه، به جای فاصله ی اقلیدسی از معیار شباهت Jensen^۱ برای سنجش میزان شباهت بین بردار مشخصه ی داده ها استفاده شده است [۷۱]. این معیار، ابزار مناسبی است برای ارزیابی شباهت در سری های زمانی نظیر این تحقیق که دارای طول یکسان اما دامنه های متفاوتی هستند.

^۱ برای اطلاع از سایر معیارهای شباهت به پیوست ۱ مراجعه کنید.

فرمول مقایسه‌ی دو بردار هم اندازه‌ی p و q به روش معیار Jensen با رابطه‌ی (۱-۳) بیان می‌شود. خروجی این فرمول، عددی بین 0 و 1 می‌باشد که اگر p و q یکسان یا کاملاً مشابه باشند؛ حاصل مقایسه عدد 0 خواهد شد:

$$\frac{1}{2} \sum_{i=1}^k \left\{ p'_i \log_2 p'_i + q'_i \log_2 q'_i - (p'_i + q'_i) \log_2 \left(\frac{p'_i + q'_i}{2} \right) \right\}$$

$$\text{where } p'_i = p_i / \sum_{i=1}^k p_i, \quad q'_i = q_i / \sum_{i=1}^k q_i, \quad (1-3)$$

برای ارزیابی عملکرد معیار شباهت Jensen، ده سری زمانی مربوط به مشخصه‌های استخراج شده از ده دسته ارقام متفاوت صفر تا ۹ و ۲۶ حرف از A تا Z مورد مقایسه قرار گرفته‌اند. هریک از این سری‌های زمانی با میانگین‌گیری از بردار مشخصه‌ی ده نمونه‌ی مختلف از هر رقم با فونت و حالات متفاوت به دست آمده‌اند. نتایج این ارزیابی در جداول ۱-۳ و ۲-۳ آمده است. نتایج این جدول نشان می‌دهد که معیار Jensen بردارهای مشخصه‌ی استخراج شده از ارقام و حروف یک دسته را کاملاً مشابه می‌داند (به علت آنکه عدد صفر را نتیجه می‌دهد) و در مقابل بردارهای مشخصه‌ی ارقام و حروف از دو دسته‌ی غیر یکسان را غیرمشابه می‌داند (عددی بزرگتر از ۰,۰۲ برای ارقام و عددی بزرگتر از ۰,۰۵ برای حروف ارائه می‌دهد).

۲-۴-۲- شبکه عصبی نگاشت خودسازمانده

در یک شبکه‌ی خودسازمانده که با SOM (Self Organization Map) یا برخی مواقع به صورت SOFM (Self Organization Feature Map) نشان داده می‌شود، واحدهای پردازشگر در نرون‌های یک شبکه یک بعدی، دو بعدی یا بیشتر قرار داده می‌شوند. واحدها، در یک فرآیند یادگیری رقابتی نسبت به الگوهای ورودی منظم می‌شوند. محل واحدهای تنظیم شده در شبکه به گونه‌ای نظم می‌یابد که برای ویژگی‌های ورودی، یک دستگاه مختصات معنی‌دار روی شبکه ایجاد شود. لذا، یک نقشه‌ی خودسازمانده، یک نقشه‌ی توپوگرافیک از الگوهای ورودی را تشکیل می‌دهد که در آن، محل قرار گرفتن واحدها، متناظر ویژگی‌های ذاتی الگوهای ورودی است. شکل ۳-۱۷، یک نمونه شبکه عصبی SOM را نشان می‌دهد.

جدول ۳-۱: استفاده از معیار شباهت Jensen برای ارزیابی شباهت بین بردارهای مشخصه‌ای استخراج شده از ارقام مختلف.

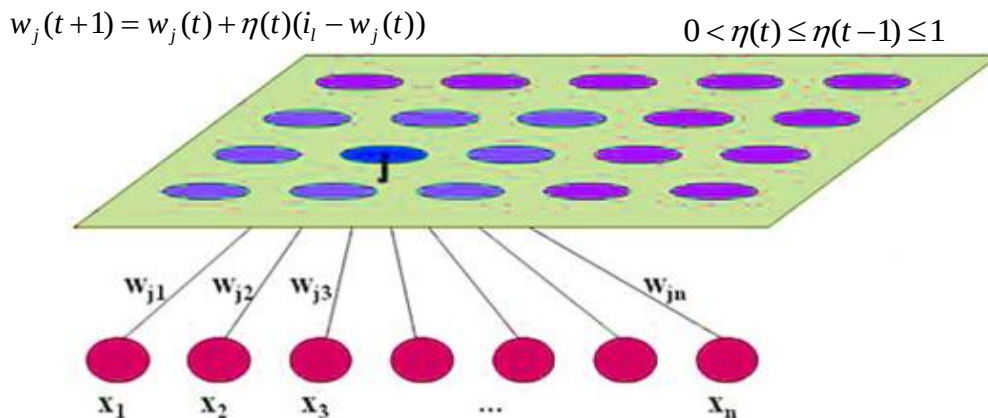
صفر	یک	دو	سه	چهار	پنج	شش	هفت	هشت	نه
صفر	۰	۰,۱۸۸۸	۰,۱۶۸۴	۰,۰۹۱۶	۰,۲۳۳۰	۰,۱۱۷۳	۰,۰۶۶۹	۰,۲۴۶۳	۰,۰۳۱۸
یک	۰	۰,۱۹۲۴	۰,۱۲۳۱	۰,۱۱۶۱	۰,۲۰۳۳	۰,۲۰۹۱	۰,۲۱۶۶	۰,۱۸۳۴	۰,۱۹۶۳
دو	۰	۰	۰,۰۴۹۵	۰,۲۰۶۷	۰,۰۵۰۳	۰,۱۱۰۸	۰,۰۸۲۹	۰,۱۲۳۳	۰,۱۱۶۹
سه	۰	۰	۰	۰,۱۶۹۴	۰,۰۳۸۶	۰,۰۷۲۲	۰,۱۰۷۸	۰,۰۶۲۹	۰,۰۶۵۲
چهار	۰	۰	۰	۰	۰,۲۱۷۳	۰,۱۶۳۰	۰,۲۰۵۷	۰,۱۷۴۱	۰,۱۹۶۷
پنج	۰	۰	۰	۰	۰	۰,۰۶۹۲	۰,۰۸۶۳	۰,۰۷۵۸	۰,۰۵۸۳
شش	۰	۰	۰	۰	۰	۰	۰,۱۷۹۱	۰,۰۲۴۰	۰,۰۴۹۸
هفت	۰	۰	۰	۰	۰	۰	۰	۰,۱۹۰۶	۰,۱۲۱۹
هشت	۰	۰	۰	۰	۰	۰	۰	۰	۰,۰۴۷۵
نه	۰	۰	۰	۰	۰	۰	۰	۰	۰

یادگیری رقابتی که در این شبکه به کار گرفته می‌شود، بدین صورت است که در هر قدم یادگیری، واحدها، با یکدیگر به رقابت می‌پردازند. در پایان یک مرحله‌ی رقابت، تنها یک واحد برنده می‌شود که وزن‌های آن نسبت به وزن‌های سایر واحدها به شکل متفاوتی تغییر داده می‌شود. الگوریتم کلی SOM شامل مراحل زیر است:

- انتخاب توپولوژی لایه خروجی شبکه.
 - مقدار دهی اولیه‌ی فاصله همسایگی جاری به یک مقدار مثبت.
- مقدار دهی اولیه‌ی وزن‌ها با مقادیر تصادفی کوچک.
- $t=1$
- تا زمانی که از کران شمارش تجاوز نکرده است کارهای زیر انجام می‌گیرد:
 - یک نمونه ورودی i_l انتخاب می‌شود.
 - توان دوم فاصله i_l از بردار وزن‌ها (فاصله اقلیدسی) برای بردار خروجی از فرمول زیر به دست می‌آید:

$$\sum_{k=1}^n (i_{l,k} - w_{j,k}(t))^2$$
 - Node خروجی ای (j^*) که دارای کمترین مقدار محاسبه شده در مرحله‌ی قبل است، انتخاب می‌شود. این نود، نود برنده است.

○ در مرحله‌ی آخر نیز پس از اعمال ورودی X ، بردارهای وزن نرون برنده و تمامی نرون‌های همجوار (همجوار به شعاع d) به سمت بردار ورودی X ، به روزرسانی می‌شوند.



شکل ۳-۱۷: یک شبکه SOM. X_i ها ورودی، w_{ji} ها وزن‌ها و z واحد برنده هستند.

○ افزودن t

• پایان حلقه

این شبکه، از یک الگوریتم یادگیری بدون مربی بهره می‌گیرد و ابزار قدرتمندی است که در کاربردهای زیادی مانند داده‌کاوی و کلاس‌بندی داده‌ها به کار گرفته می‌شود [۷۲]. این نوع شبکه عصبی بیشترین شباهت را با پدیده‌های زیستی دارد و بر اساس نظریه‌های “Winner takes all (WTA)” یعنی برنده همه چیز را می‌برد و “Winner takes most (WTM)” یعنی برنده بیشتر می‌برد، بنا نهاده شده است.

بنابراین، به طور خلاصه می‌توان نحوه عملکرد SOM را این‌گونه بیان کرد: اگر بخواهیم داده‌ها را به n دسته تقسیم کنیم، n نرون را برای لایه خروجی در نظر می‌گیریم. بنابراین، ماتریس وزن‌ها دارای ابعاد $n \times m$ خواهد بود که m نیز اندازه‌ی عناصر هر یک از نمونه‌های ورودی است که مقداردهی اولیه وزن‌ها با مقادیر تصادفی کوچک صورت می‌گیرد. سپس هر داده‌ی ورودی را با هر یک از سطری‌های ماتریس وزن مقایسه (با استفاده از فاصله اقلیدسی یا معیارهای شباهت) و به هر کدام که نزدیکتر بود نرون متناظر با آن سطر را برنده اعلام کرده و آن سطر از ماتریس وزن را به‌روزرسانی می‌کنیم ((در بعضی انواع

طراحی این نوع از شبکه عصبی گاهی وزن‌های متصل به نرون همسایه با نرون برنده نیز به‌روز رسانی می‌شود)).

در این پایان‌نامه از شبکه‌ی عصبی SOM برای دسته‌بندی ارقام و حروف استفاده می‌شود. اندازه لایه‌ی ورودی برابر با اندازه‌ی بردار مشخصه می‌باشد که برای ارقام برابر ۹۰ و برای حروف برابر ۱۶۵ است. در شبکه عصبی تشخیص ارقام، از آنجایی که هر رقم در یک کلاس مجزا قرار می‌گیرد؛ این شبکه دارای ده نرون در لایه‌ی خروجی می‌باشد. شبکه عصبی حروف، ۲۶ نرون در لایه‌ی خروجی دارد که به تعداد حروف انگلیسی است.

در مرحله‌ی آموزش، وزن دهی اولیه‌ی شبکه با اعداد تصادفی کوچکتر از واحد انجام گرفته است. در این شبکه، برخلاف شبکه‌های SOM معمول که از فاصله‌ی اقلیدسی استفاده می‌کنند، از معیار شباهت Jensen برای تعیین نرون برنده استفاده شده است. آموزش شبکه بر اساس قاعده‌ی آموزشی شبکه‌های SOM تکرار می‌شود و در هر بار تکرار، وزن واحد برنده به روزرسانی می‌گردد تا مناسب‌ترین نماینده‌ی هریک از ارقام تعیین شود.

در این تحقیق، برای آموزش شبکه عصبی ارقام، تنها یک نمونه از هر رقم به شبکه داده می‌شود. با اتمام آموزش شبکه، وزن هر یک از نرون‌ها در لایه‌ی خروجی نماینده‌ی یکی از ارقام خواهد بود. با اعمال نمونه‌های آزمایشی به شبکه شبیه‌ترین وزن نرون به آن نمونه انتخاب می‌شود و بر این اساس عمل دسته‌بندی انجام می‌گیرد.

به منظور آموزش شبکه عصبی حروف، از هر کدام از کلاس‌های حروف تنها دو نمونه جهت آموزش داده شده است؛ این عمل تاکید می‌کند که نیازی به حضور نمونه‌ای از هر یک از فونت‌ها برای آموزش نیست تا در هنگام تست یک نمونه، سیستم قبلاً نمونه‌ای از آن را مشاهده کرده باشد. در واقع، معیار شباهت با تشخیص مشابه بودن حروف یک دسته به یکدیگر با فونت‌های مختلف، سبب می‌شود تا اگر نمونه‌هایی با فونت‌های ناشناخته نیز به سیستم داده شود، در شناسایی آنها موفق عمل کند.

به این دلیل دو نمونه از هر کدام از حروف برای آموزش داده می‌شود که برخی از حروف مانند M، E، F و G به دو صورت متفاوت (شکل ۳-۱۸) نوشته می‌شوند که در صورت عدم آموزش این دو نمونه، سیستم در تشخیص موفق عمل نخواهد کرد.

E E F F G G M M

شکل ۳-۱۸: نمونه‌های مختلف نوشتاری برخی از حروف.

۳-۲-۵- انتخاب ویژگی‌ها با الگوریتم وراثتی

ویژگی‌های ذکر شده در بخش ۳-۲-۳، یک بردار ویژگی با بعد بالا به دست می‌دهد. (۹۰ ویژگی برای ارقام و ۱۶۵ ویژگی برای حروف) یک راه برای کاهش بعد، حذف همبستگی بین ابعاد به کمک روش‌هایی مثل PCA و ICA است که باعث کم شدن هزینه‌ی طبقه‌بندی می‌شود. اما این تکنیک‌ها، تاثیری در تعداد ویژگی‌هایی که لازم است از تصاویر استخراج شود، ندارد. در این تحقیق، کاهش بعد به کمک الگوریتم وراثتی پیشنهاد شده است که علاوه بر کاهش بعد، تعداد ویژگی‌های لازم را نیز کاهش می‌دهد.

الگوریتم وراثتی، یکی از بهترین روش‌های موجود برای جستجوی بهینه است [۷۳]. این الگوریتم برای انتخاب ویژگی‌های برتر، ابزار مفید و کارآمدی است. انتخاب ویژگی با الگوریتم وراثتی در مرجع [۷۴] ارائه شده است. در این روش از الگوریتم وراثتی برای یافتن بردار باینری بهینه‌ای استفاده می‌شود که هر یک از بیت‌های آن متناظر یک ویژگی است. اگر i امین بیت این بردار یک باشد، i امین ویژگی در طبقه بندی شرکت کرده و در صورتی که صفر باشد، ویژگی متناظر آن حذف می‌شود. در این روش، هر یک از زیرمجموعه‌های ویژگی با توجه به نرخ طبقه بندی درست آن روی مجموعه آزمایش با استفاده از طبقه‌بند k همسایه نزدیکتر ارزیابی می‌شوند [۷۵].

در یک الگوریتم وراثتی، یک جمعیت از افراد در محیط بقا می‌یابند. افرادی با قابلیت‌های بالاتر، شانس ازدواج و تولید مثل بیشتری را خواهند یافت. بنابراین بعد از چند نسل، جمعیتی با کارایی بهتر به وجود می‌آید. در الگوریتم وراثتی، هر فرد از جمعیت به صورت یک کروموزوم معرفی می‌شود. کروموزوم‌ها در طول چندین نسل کامل‌تر می‌شوند. در هر نسل، کروموزوم‌ها ارزیابی می‌گردند و متناسب با ارزش خود، امکان بقا و تکثیر می‌یابند. تولید نسل در بحث الگوریتم وراثتی با عملگرهای انتخاب، جهش^۱ و ترکیب^۲ صورت می‌گیرد. والدین برتر بر اساس یک تابع برازندگی^۳ انتخاب می‌شوند. در عمل ترکیب،

^۱ mutation

^۲ crossover

^۳ Fitness function

کروموزوم‌هایی که برازندگی آنها بیشتر باشد، شانس بیشتری برای انتخاب شدن دارند. این عمل به صورت اتفاقی ولی بر اساس مقدار تابع برازندگی صورت می‌گیرد.

در نتیجه، مراحل انتخاب مشخصه‌ها را می‌توان به صورت زیر بیان کرد:

۱. تشکیل جمعیت اولیه.
 ۲. انتخاب کروموزوم‌های با مقدار شایستگی بالاتر برای ایجاد فرزندان.
 ۳. انجام ترکیب.
 ۴. انجام جهش.
 ۵. جابجایی کروموزوم‌های حاصل در مرحله ۴ به جای والدین در صورت داشتن شایستگی بیشتر از والدین.
 ۶. رفتن به گام ۲ تا برقراری شرط خاتمه.
- شرط خاتمه، اتمام تعداد ثابتی از تکرارهاست که ما تکرار آن را ۲۰ فرض کرده‌ایم. طول هر کروموزوم به اندازه‌ی تعداد ویژگی‌هاست که برای ارقام، طول کروموزوم برابر ۹۰ و برای حروف برابر ۱۶۵ است. جمعیت اولیه به تعداد ۲۰ نمونه به صورت اعداد رندم ۰ و ۱ ایجاد می‌شود. تابع برازندگی مورد استفاده، بررسی صحت دسته‌بندی است. در واقع هر بار جمعیت موجود را به شبکه‌ی SOM داده و دقت دسته‌بندی را با این کروموزوم‌ها محاسبه می‌کنیم. در هر تکرار، ۴ نمونه از بهترین کروموزوم‌ها (کروموزوم‌های با شایستگی بالاتر یا به عبارتی دارای دقت بیشتر) را انتخاب کرده و با آنها عمل ترکیب و جهش را انجام می‌دهیم و مجدداً میزان شایستگی آنها را می‌سنجیم. در صورت بیشتر بودن این میزان، این کروموزوم‌ها، جایگزین والدین خود می‌شوند و این روند در تکرارهای بعد ادامه می‌یابد تا زمان خاتمه فرا برسد.

با اعمال الگوریتم وراثتی بر روی بردار ویژگی ۹۰ تایی به دست آمده برای ارقام، توانستیم طول این بردار را به ۴۰ ویژگی کاهش دهیم. این کاهش ویژگی نشان می‌دهد که جمع سطرها و ستون‌های حذف شده از مجموعه‌ی ویژگی‌ها در تصاویر مختلف متمایز نبوده و توانایی جدا کردن ارقام از یکدیگر را نداشته‌اند.

با اعمال این الگوریتم بر بردار ویژگی ۱۶۵ تایی به دست آمده برای حروف بزرگ، تنها ۳۰ ویژگی به عنوان ویژگی‌های مناسب جهت دسته‌بندی حروف شناخته شدند. این ویژگی‌ها نشان می‌دهند که جمع تمام سطرها و ستون‌ها برای متمایز کردن حروف نیاز نبوده و تنها ۳۰ نمونه از آنها کافی هستند. هم‌چنین، برای شناسایی حروف کوچک، ۴۰ ویژگی از مجموعه ویژگی‌های ۱۶۵ تایی انتخاب شده‌اند که بیانگر این است که برای شناسایی حروف کوچک، تعداد ویژگی‌های تمییزدهنده‌ی بین حروف، از ویژگی‌های حروف بزرگ بیشتر هستند.

۳-۳- نتیجه‌گیری

در این فصل، سیستم پیشنهادی برای شناسایی ارقام و حروف تایی انگلیسی بدون محدودیت در اندازه و نوع فونت را شرح دادیم. این سیستم، با استخراج ویژگی‌های ساده‌ای - مجموع یک‌های تک تک سطرها و ستون‌ها- و شبکه عصبی نگاشت خودسازمانده بدون ساختار پیچیده، قادر است ارقام و حروف را با دقت بالایی بازشناسی کند. این شبکه عصبی به جای استفاده از فاصله‌ی اقلیدسی برای مقایسه‌ی بین نمونه‌ها و وزن‌های برنده، از معیار شباهت استفاده می‌کند. هم‌چنین، برای انتخاب ویژگی‌ها از بردار ویژگی به دست آمده، از الگوریتم وراثتی استفاده شد. در فصل بعدی نتایج حاصل از اعمال روش پیشنهادی را بر روی پایگاه داده‌ی معرفی شده بیان خواهیم کرد.

جدول ۳-۲: استفاده از معیار شباهت Jensen برای ارزیابی شباهت بین بردارهای مشخصه‌ی استخراج شده از حروف مختلف.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	0	0.10	0.24	0.17	0.16	0.19	0.18	0.12	0.24	0.25	0.20	0.19	0.13	0.16	0.20	0.20	0.16	0.09	0.12	0.25	0.18	0.30	0.20	0.23	0.28	0.17
B		0	0.20	0.05	0.06	0.09	0.15	0.07	0.20	0.21	0.13	0.12	0.09	0.08	0.14	0.10	0.15	0.04	0.06	0.18	0.07	0.16	0.14	0.16	0.15	0.14
C			0	0.20	0.15	0.25	0.03	0.26	0.26	0.24	0.23	0.22	0.25	0.23	0.08	0.28	0.09	0.22	0.15	0.22	0.15	0.23	0.21	0.27	0.27	0.24
D				0	0.12	0.09	0.16	0.14	0.19	0.21	0.16	0.10	0.14	0.12	0.10	0.05	0.18	0.10	0.11	0.17	0.12	0.15	0.17	0.20	0.16	0.17
E					0	0.11	0.13	0.10	0.15	0.16	0.08	0.12	0.16	0.10	0.20	0.17	0.13	0.06	0.06	0.14	0.07	0.14	0.13	0.12	0.15	0.11
F						0	0.20	0.14	0.11	0.19	0.13	0.11	0.14	0.13	0.24	0.04	0.16	0.09	0.12	0.10	0.14	0.13	0.19	0.13	0.10	0.15
G							0	0.20	0.21	0.19	0.17	0.18	0.19	0.17	0.07	0.24	0.07	0.17	0.10	0.19	0.13	0.20	0.13	0.20	0.22	0.19
H								0	0.28	0.26	0.10	0.20	0.06	0.06	0.20	0.19	0.17	0.05	0.09	0.23	0.05	0.15	0.13	0.13	0.16	0.15
I									0	0.08	0.18	0.12	0.27	0.23	0.32	0.14	0.17	0.19	0.19	0.05	0.26	0.20	0.19	0.17	0.16	0.16
J										0	0.20	0.20	0.29	0.21	0.28	0.22	0.17	0.21	0.17	0.11	0.20	0.21	0.14	0.19	0.19	0.19
K											0	0.14	0.15	0.09	0.25	0.19	0.17	0.10	0.10	0.15	0.11	0.16	0.10	0.04	0.18	0.10
L												0	0.20	0.18	0.23	0.13	0.21	0.13	0.15	0.17	0.21	0.21	0.19	0.16	0.24	0.11
M													0	0.11	0.18	0.17	0.17	0.10	0.11	0.21	0.11	0.19	0.15	0.17	0.13	0.17
N														0	0.19	0.16	0.16	0.06	0.08	0.19	0.08	0.19	0.10	0.10	0.17	0.15
O															0	0.20	0.13	0.20	0.14	0.26	0.15	0.24	0.19	0.31	0.26	0.27
P																0	0.20	0.14	0.16	0.12	0.19	0.17	0.21	0.20	0.14	0.19
Q																	0	0.14	0.13	0.14	0.13	0.16	0.14	0.19	0.16	0.18
R																		0	0.08	0.17	0.09	0.19	0.14	0.12	0.18	0.11
S																			0	0.16	0.08	0.17	0.12	0.13	0.14	0.14
T																				0	0.20	0.16	0.17	0.14	0.10	0.14
U																					0	0.11	0.11	0.16	0.14	0.17
V																						0	0.16	0.23	0.09	0.24
W																							0	0.14	0.20	0.15
X																								0	0.19	0.08
Y																									0	0.24
Z																										0

فصل چهارم

ارزیابی عملکرد و تحلیل نتایج

۱-۴- مقدمه

در این فصل، ابتدا معیارهای دقت و ارزیابی عملکرد را توضیح داده و سپس به بیان نتایج حاصل از دسته‌بندی ارقام، حروف بزرگ و کوچک و در نهایت ارقام فارسی می‌پردازیم.

برای معرفی معیار دقت و دیگر معیارهای کارایی، لازم است بدانیم که با فرض داشتن یک مجموعه با دو کلاس بله و خیر، یک پیش‌بینی می‌تواند دارای ۴ حالت باشد (جدول ۱-۴).

جدول ۱-۴: حالات مختلف یک پیش‌بینی دو کلاسه.

کلاس پیش‌بینی شده			
خیر	بله		
منفی اشتباه (FN)	مثبت واقعی (TP)	بله	کلاس درست
منفی واقعی (TN)	مثبت اشتباه (FP)	خیر	

مثبت واقعی (TP) دسته‌بندی صحیح می‌باشد. مثبت اشتباه (FP) و منفی اشتباه (FN) زمانی رخ می‌دهند که نمونه‌ای که واقعاً منفی است، مثبت و نمونه‌ای که به طور واقعی مثبت می‌باشد، منفی پیش‌بینی گردد. بنابراین، دقت یا نسبت موفقیت کلی، نسبت تعداد دسته‌بندی‌های درست به کل دسته‌بندی‌ها است (معادله ۱-۴).

$$AC = \frac{TP}{TP + TN + FP + FN} \quad (1-4)$$

هم‌چنین برای ارزیابی یادگیرنده از پارامترهای دیگری مانند فراخوانی^۱، دقت^۲ یا اندازه-اف^۳ نیز استفاده می‌شود [۲۲]. پارامتر فراخوانی بیان می‌کند که یادگیرنده چه نسبتی از کلاس‌های مثبت را درست پیش‌بینی می‌کند (فرمول ۲-۴).

¹ Recall

² Precision

³ F-Measure

$$R = \frac{TP}{TP + FN} \quad (2-4)$$

پارامتر دقت بیان می‌کند که چند درصد از رکوردهایی که یادگیرنده مثبت معرفی کرده واقعاً مثبت هستند. این مفهوم در فرمول (3-4) نشان داده شده است.

$$PR = \frac{TP}{TP + FP} \quad (3-4)$$

اندازه-ف نیز ترکیب و میانگین هارمونیک این دو پارامتر می‌باشد که با استفاده از فرمول (4-4) بدست می‌آید.

$$FM = 2 * \frac{PR * R}{PR + R} \quad (4-4)$$

اندازه-ف کلی^۱ در رابطه‌ی (5-4) نیز، از مجموع نسبت حاصلضرب اندازه-ف در تعداد TP‌های هر کلاس به تعداد کل داده‌ها به دست می‌آید. (n تعداد کل داده‌هاست)

$$TFM = \sum \frac{TP}{n} * (FM) \quad (5-4)$$

۲-۴- نتایج

۲-۴-۱- نتایج حاصل از دسته‌بندی ارقام انگلیسی

حال برای ارزیابی عملکرد روش پیشنهادی برای دسته‌بندی ارقام، از پایگاه داده‌ی معرفی شده در بخش ۱-۲-۳ استفاده می‌کنیم. در این آزمایش، عملکرد روش پیشنهادی در تشخیص ارقام با فونت‌ها و اندازه‌های معرفی شده، ارزیابی می‌شود. با اعمال روش پیشنهادی، ارقام موجود در پایگاه داده‌ی معرفی شده را تشخیص و دسته‌بندی می‌کنیم. جدول ۲-۴، دقت روش پیشنهادی را بر اساس روابط (2-4)، (4-4) و (3-4) نشان می‌دهد.

¹ Total F-Measure

جدول ۴-۲: دقت روش پیشنهادی بر روی داده‌های ارقام.

	صفر	یک	دو	سه	چهار	پنج	شش	هفت	هشت	نه
PR	۰,۹۷	۰,۹۸	۱	۱	۱	۱	۱	۱	۰,۹۹	۱
R	۱	۱	۱	۱	۱	۱	۰,۹۷	۰,۹۸	۱	۱
FM	۰,۹۸	۰,۹۹	۱	۱	۱	۱	۰,۹۸	۰,۹۹	۰,۹۹	۱

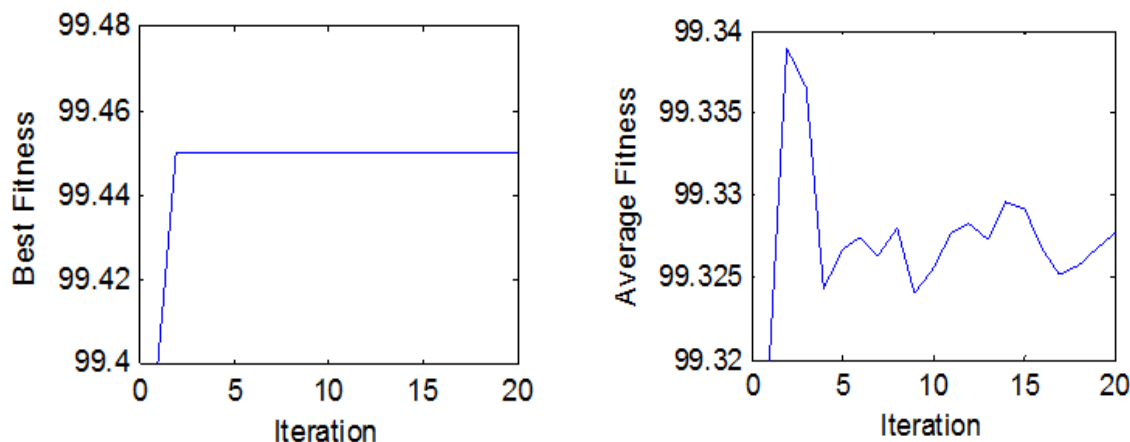
مقادیر TFM و AC حاصل از ارزیابی روش پیشنهادی به ترتیب برابر ۰,۹۹۱۱ و ۹۹,۵۵٪ هستند. همچنین، جدول ۴-۳، دقت دسته‌بندی هریک از ارقام صفر تا نه را نشان می‌دهد. همانطور که از این جدول برمی‌آید، روش پیشنهادی در شناسایی رقم ۶ ضعیف عمل کرده است؛ این امر می‌تواند به سبب شکل‌های بسیار متفاوت این رقم در فونت‌های مختلف باشد که معیار شباهت قادر نبوده شباهت بین آنها را تشخیص دهد.

جدول ۴-۳: نتیجه‌ی دسته‌بندی ارقام با روش پیشنهادی.

ارقام	تعداد تشخیص درست در نمونه‌ها	دقت
صفر	۲۰۰	٪۱۰۰
یک	۲۰۰	٪۱۰۰
دو	۲۰۰	٪۱۰۰
سه	۲۰۰	٪۱۰۰
چهار	۲۰۰	٪۱۰۰
پنج	۲۰۰	٪۱۰۰
شش	۱۹۴	٪۹۷
هفت	۱۹۷	٪۹۸,۵
هشت	۲۰۰	٪۱۰۰
نه	۲۰۰	٪۱۰۰
دقت کل		٪۹۹,۵۵

همانطور که در بخش ۳-۲-۵ شرح داده شد، توانستیم با استفاده از الگوریتم وراثتی تعداد ویژگی‌های استفاده شده برای تشخیص را کاهش داده و از ۹۰ ویژگی به ۴۰ ویژگی برسیم. شکل ۴-۱، نمودارهای بهترین مقدار برازندگی و متوسط آن را در ۲۰ تکرار الگوریتم وراثتی نشان می‌دهد. متأسفانه با

کاهش ویژگی‌ها از ۹۰ به ۴۰، حدود ۱۰٪ از میزان دقت شناسایی کم شد؛ اما شناسایی هر رقم در مرحله‌ی تست با بردار ویژگی ۹۰ تایی در سیستمی با پردازنده‌ی دوهسته‌ای 2.10 GHz و حافظه‌ی رم 4GB، ۰.۰۲۹ ثانیه و پس از اعمال الگوریتم وراثتی ۰.۰۱۱ ثانیه زمان می‌برد که کاهش دقت، به این افزایش سرعت و کم کردن محاسبات در مرحله‌ی تست می‌ارزد.



شکل ۴-۱: نمودارهای بهترین برازندگی و متوسط برازندگی در ۲۰ تکرار الگوریتم برای کاهش ویژگی‌های سیستم شناسایی ارقام.

جدول ۴-۴، روش پیشنهادی را با سایر روش‌های موجود مقایسه می‌کند. دو روش دیگر بیان شده، ارقام را در ۱۱ اندازه و ۱۷ فونتی دسته‌بندی کرده‌اند که ما نیز نمونه‌هایی با همان ۱۷ فونت و ۱۱ اندازه را در پایگاه داده‌مان داریم. بنابراین می‌توان گفت که روش ما علاوه بر آن ۱۷ فونت، نمونه‌هایی با ۱۳ فونت جدید را نسبت به آن دو روش با دقت بالایی شناسایی کرده است. همانطور که مشاهده می‌شود، تعداد کم نمونه‌های آموزشی، تعداد کم ویژگی‌های مورد استفاده، مقاوم بودن نسبت به تغییر فونت، روش کلاس‌بندی ساده و دقت دسته‌بندی بالا از مزایای روش پیشنهادی نسبت به روش‌های دیگر می‌باشد.

همانطور که در جدول ۴-۴، ذکر شده است دقت دسته‌بندی با ۱۷ فونت برابر ۹۹.۵۲٪ و دقت دسته‌بندی با ۳۰ فونت برابر ۹۹.۴۵٪ است؛ این مقادیر نشان می‌دهند که با وجود افزایش تعداد فونت‌ها، دقت تشخیص به طور قابل ملاحظه‌ای تغییر نکرده است و خطای بازشناسی تنها به میزان ۰.۰۷٪ افزایش یافته است. به عبارت دیگر، می‌توان گفت که روش پیشنهادی در مقابل تغییر فونت مقاوم است و با افزایش تعداد فونت‌ها به مقداری بیش از ۳۰، خطای تشخیص تغییر چندانی نخواهد داشت.

جدول ۴-۴: مقایسه‌ی روش پیشنهادی شناسایی ارقام با سایر روش‌های موجود در این حوزه.

تعداد نمونه‌های آموزشی و آزمایشی	روش کلاس‌بندی	تعداد فونت‌های آزمایش شده	تعداد ویژگی‌ها	دقت
Dhandra, et al.[59] ۱۶۰ آموزش- ۳۲۰۰ تست	درخت تصمیم	۱۶	۶۰	۹۹,۷۸٪
Rani et al.[61] ۹۳۵ آموزش- ۱۸۷۰ تست	ماشین بردار پشتیبان	۱۷	۲۰۰	۹۸,۱۲٪
روش پیشنهادی ۱۰ آموزش- ۱۸۷۰ تست	شبکه عصبی نگاشت خودسازمانده	۱۷	۴۰	۹۹,۵۲٪
روش پیشنهادی ۱۰ آموزش- ۱۸۷۰ تست	شبکه عصبی نگاشت خودسازمانده	۳۰	۹۰	۹۹,۵۵٪
روش پیشنهادی ۱۰ آموزش- ۲۰۰۰ تست	شبکه عصبی نگاشت خودسازمانده	۳۰	۴۰	۹۹,۴۵٪

۴-۲-۲- نتایج حاصل از دسته‌بندی حروف انگلیسی

برای دسته‌بندی حروف بزرگ، از پایگاه داده‌ی معرفی شده در بخش ۳-۲-۱ استفاده کردیم. جدول ۴-۵، ارزیابی روش پیشنهادی را بر اساس روابط (۴-۲)، (۴-۳) و (۴-۴) نشان می‌دهد.

مقادیر TFM و AC حاصل از ارزیابی روش پیشنهادی به ترتیب برابر ۰,۹۷۰۰ و ۹۸,۴۷٪ هستند. جدول ۴-۶، دقت دسته‌بندی هر یک از حروف را با روش پیشنهادی نشان می‌دهد. همانطور که مشاهده می‌شود، این روش در شناسایی حرف M ضعیف عمل کرده است. این عملکرد ضعیف، ناشی از شباهت زیاد سری زمانی این حرف به حرف H است (به شکل ۳-۱۱ توجه کنید) و در نتیجه اشتباهاً حرف M در دسته‌ی حرف H قرار گرفته است. این نشان می‌دهد که ویژگی‌های استخراجی نیاز به بازبینی

داشته و برای برخی حروف مانند این حالت، نیاز است که ویژگی‌های دیگری برای تشخیص با نرخ بالاتر به کار برد. همچنین، نمونه‌ای که به عنوان نماینده جهت آموزش به سیستم داده می‌شود، نقش تاثیرگذاری در میزان دقت حاصل از دسته‌بندی دارد؛ اگر این نماینده، متناسب با هر یک از حروف یا ارقام یک دسته و به صورت منحصر به فرد برای آن دسته و متفاوت از سایر دسته‌ها مشخص شود، عملکرد سیستم بهتر خواهد بود.

جداول ۷-۴ و ۸-۴، به ترتیب روابط (۲-۴)، (۳-۴) و (۴-۴) را با اعمال روش پیشنهادی بر پایگاه داده‌ی معرفی شده در بخش ۱-۲-۳ و دقت دسته‌بندی را برای هریک از حروف کوچک نشان می‌دهند.

همانطور که مشاهده می‌شود، روش پیشنهادی در شناسایی حروف کوچک چندان موفق عمل نکرده است. نحوه‌ی نوشتن حروف کوچک مربوط به یک دسته بسیار متنوع هستند و این تفاوت بسیار در فونت‌های مختلف، سبب شده است که معیار شباهت قادر نباشد حروف یک دسته را با وجود تفاوت‌های آنها شبیه به هم شناسایی کند؛ بنابراین، نیازمند ویژگی‌های بسیار دقیق‌تر با خاصیت متمایزکنندگی بیشتر هستیم تا بتوانیم حروف کوچک را نیز با دقت بالایی تشخیص دهیم. شکل ۲-۴، برخی از این حروف و تنوع نوشتاری آنها را نشان می‌دهد. مقدار TFM و AC شناسایی حروف کوچک به ترتیب برابر ۰,۸۶۴۵ و ۹۲,۸۷٪ می‌باشد.

a α g g l l y y

شکل ۲-۴: نمونه‌هایی از حروف a, g, l و y که به دو صورت کاملاً متفاوت در فونت‌های مختلف نوشته می‌شوند.

جدول ۴-۵: دقت روش پیشنهادی بر روی حروف بزرگ.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
PR	۱	۰,۹۹	۱	۰,۹۹	۰,۹۷	۱	۰,۹۸	۰,۸۹	۰,۹۸	۱	۰,۹۹	۱	۰,۹۹	۱	۰,۹۷	۰,۹۹	۰,۹۹	۰,۹۵	۱	۰,۹۹	۰,۹۶	۱	۱	۱	۰,۹۸	۰,۹۹
R	۰,۹۹	۰,۹۷	۰,۹۸	۰,۹۹	۰,۹۹	۱	۰,۹۹	۰,۹۸	۱	۰,۹۹	۰,۹۹	۰,۹۹	۰,۹۰	۰,۹۴	۰,۹۸	۰,۹۸	۰,۹۸	۰,۹۸	۰,۹۹	۱	۱	۰,۹۸	۱	۱	۰,۹۹	۱
FM	۰,۹۹	۰,۹۸	۰,۹۹	۰,۹۹	۰,۹۸	۱	۰,۹۸	۰,۹۳	۰,۹۹	۰,۹۹	۰,۹۹	۰,۹۹	۰,۹۴	۰,۹۷	۰,۹۸	۰,۹۹	۰,۹۸	۰,۹۶	۰,۹۹	۰,۹۹	۰,۹۸	۰,۹۹	۱	۱	۰,۹۹	۰,۹۹

جدول ۴-۶: نتیجه دسته‌بندی روش پیشنهادی بر روی حروف بزرگ.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
۲۳۵	۲۳۱	۲۳۳	۲۳۵	۲۳۴	۲۳۷	۲۳۵	۲۳۳	۲۳۷	۲۳۴	۲۳۵	۲۳۶	۲۱۳	۲۲۳	۲۳۲	۲۳۲	۲۳۲	۲۳۳	۲۳۵	۲۴۷	۲۳۷	۲۳۲	۲۴۷	۲۳۷	۲۳۶	۲۳۷	تعداد تشخیص درست در نمونه‌ها
۹۹	۹۷,۴۷	۹۸,۳۱	۹۹,۱۶	۹۹	۱۰۰	۹۹,۱۶	۹۸,۳۱	۱۰۰	۹۹	۹۹	۹۹	۹۰	۹۴,۰۹	۹۸	۹۸	۹۸	۹۸	۹۹	۱۰۰	۱۰۰	۹۸	۱۰۰	۱۰۰	۹۹,۵۷	۱۰۰	درصد دقت دسته بندی
																						دقت کل: ۹۸,۴۷٪				

جدول ۴-۷: دقت روش پیشنهادی بر روی حروف کوچک.

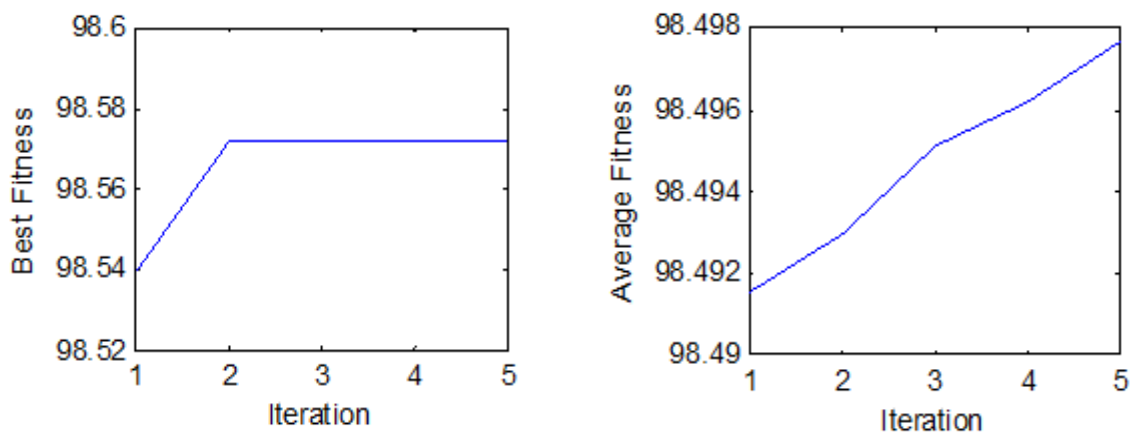
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
PR	۰,۸۶	۱	۰,۹۰	۰,۹۶	۰,۹۸	۰,۸۹	۰,۷۶	۰,۹۴	۰,۹۸	۰,۹۵	۰,۹۶	۰,۷۸	۱	۰,۸۹	۰,۹۰	۰,۹۵	۰,۹۹	۰,۹۵	۰,۹۵	۰,۸۷	۰,۹۰	۱	۱	۱	۰,۹۸	۰,۸۷
R	۰,۹۲	۰,۹۴	۰,۹۴	۱	۰,۸۵	۰,۹۵	۰,۸۹	۰,۹۴	۰,۹۰	۰,۹۹	۰,۹۰	۰,۸۹	۰,۹۵	۰,۹۰	۰,۸۶	۰,۹۸	۰,۹۵	۰,۹۰	۰,۹۸	۰,۸۸	۰,۸۵	۰,۸۶	۱	۰,۹۰	۰,۹۵	۱
FM	۰,۸۹	۰,۹۷	۰,۹۲	۰,۹۸	۰,۹۱	۰,۹۲	۰,۸۲	۰,۹۴	۰,۹۴	۰,۹۷	۰,۹۳	۰,۸۳	۰,۹۸	۰,۹۰	۰,۸۸	۰,۹۷	۰,۹۷	۰,۹۲	۰,۹۷	۰,۸۸	۰,۸۸	۰,۹۳	۱	۰,۹۵	۰,۹۶	۰,۹۳

جدول ۴-۷: نتیجه دسته‌بندی روش پیشنهادی بر روی حروف کوچک.

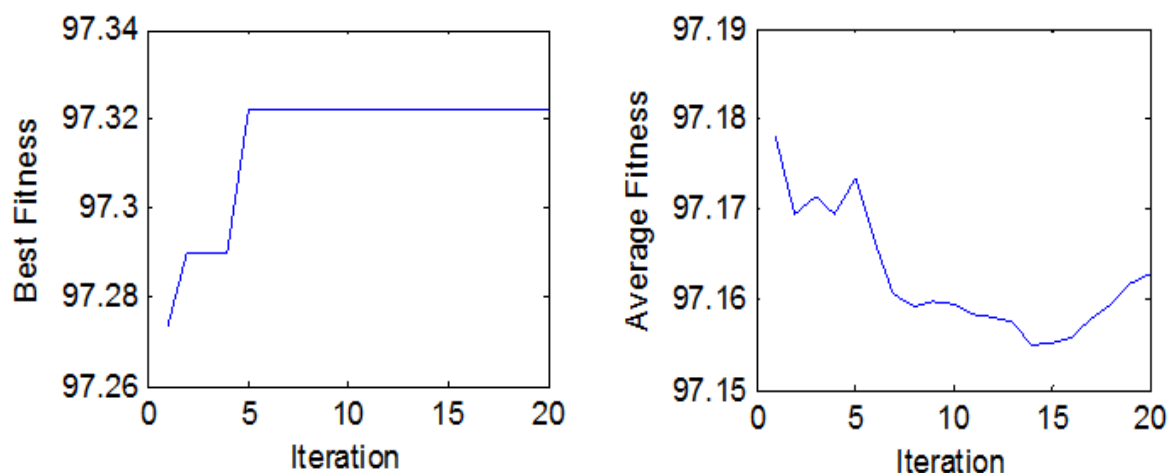
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	
۲۱۸	۲۲۲	۲۲۲	۲۳۷	۲۰۱	۲۲۵	۲۱۰	۲۲۲	۲۱۳	۲۳۵	۲۱۳	۲۱۰	۲۲۵	۲۱۳	۲۰۳	۲۳۲	۲۲۵	۲۱۳	۲۳۲	۲۰۸	۲۰۱	۲۰۳	۲۳۷	۲۱۳	۲۲۵	۲۳۷	تعداد تشخیص درست نمونه‌ها
۹۲	۹۴	۹۴	۱۰۰	۸۵	۹۵	۸۹	۹۴	۹۰	۹۹	۹۰	۸۹	۹۵	۹۰	۸۶	۹۸	۹۵	۹۰	۹۸	۸۸	۸۵	۸۶	۱۰۰	۹۰	۹۵	۱۰۰	درصد دقت دسته بندی
																						دقت کل: ۹۲,۸۷٪				

همانطور که در بخش ۳-۲-۵ بیان شد، برای انتخاب ویژگی‌ها و کاهش ابعاد بردار ویژگی از الگوریتم وراثتی استفاده کردیم. با استفاده از این الگوریتم، ابتدا ۱۶۵ ویژگی را به ۸۰ ویژگی کاهش دادیم. این کاهش ویژگی، سبب افزایش دقت از ۹۸,۴۳٪ به ۹۸,۵۶٪ شد. شکل ۴-۳، نمودار تغییر مقدار بهترین برازندگی و متوسط آن را نشان می‌دهد. مجدداً، الگوریتم وراثتی را بر روی این بردار ویژگی ۸۰ تایی اعمال کردیم و توانستیم از ۸۰ ویژگی تنها ۳۰ ویژگی را برای دسته‌بندی به کار ببریم. شکل ۴-۴، نشان‌دهنده نمودار تغییر مقدار بهترین برازندگی و متوسط میزان آن در ۲۰ تکرار الگوریتم وراثتی است. همانطور که مشاهده می‌شود با کاهش ویژگی‌ها از ۸۰ به ۳۰، حدود ۱٪ از دقت شناسایی کاسته شده است؛ اما پیش از اعمال الگوریتم وراثتی در سیستمی با پردازنده‌ی دوهسته‌ای 2.10 GHz و حافظه‌ی رم 4GB، شناسایی هر حرف (بردار ویژگی ۱۶۵ تایی) در مرحله‌ی تست ۰,۰۴۲ ثانیه به طول انجامید در حالی که پس از کاهش ویژگی‌ها، این زمان به ۰,۰۱۳ ثانیه کاهش یافت. به جهت این افزایش سرعت شناسایی در مرحله‌ی تست، کاهش دقت در ازای این کاهش ویژگی اهمیت دارد. هرچند که با تغییر مقدار ضریب یادگیری از ۰,۰۳ به ۰,۰۵ در هنگام آموزش، توانستیم با ۳۰ ویژگی، دقت را از مقدار ۹۷,۳۲٪ به ۹۷,۷۳٪ افزایش دهیم.

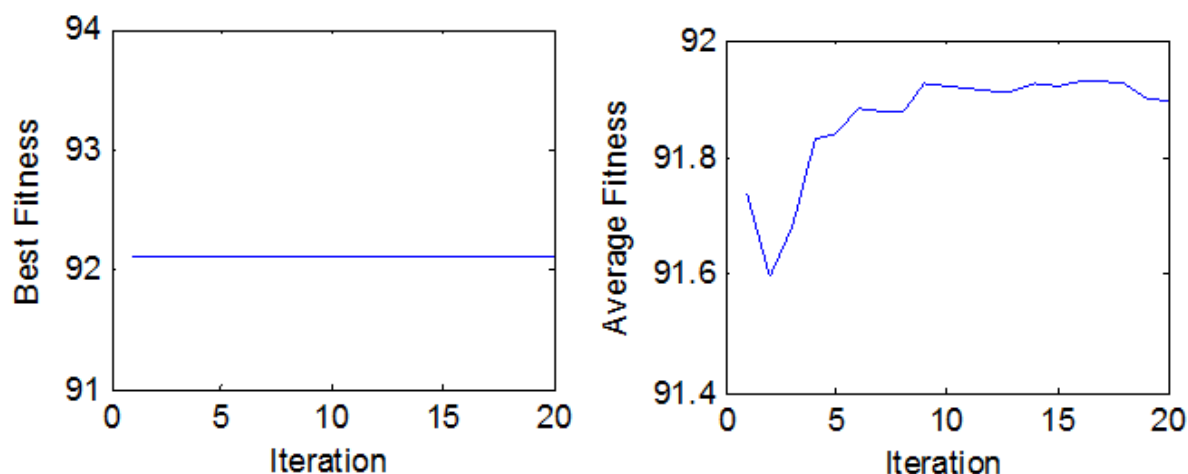
با اعمال الگوریتم وراثتی بر روی داده‌های حروف کوچک، ویژگی‌ها از ۱۶۵ به ۴۰ ویژگی کاهش یافتند. شکل ۴-۴، نمودارهای بهترین و متوسط مقدار برازندگی را برای این داده‌ها نمایش می‌دهند. با این کاهش ویژگی از دقت ۹۲,۸۷٪ به ۹۲,۱۰٪ رسیدیم که حدود ۰,۷٪ کاهش دقت داشتیم؛ اما همانطور که پیش از این اشاره شد، این کاهش دقت به افزایش سرعت مرحله‌ی تست می‌ارزد.



شکل ۴-۳: نمودارهای بهترین برازندگی و متوسط برازندگی در ۵ تکرار الگوریتم برای کاهش ویژگی حروف بزرگ از ۱۶۵ به ۸۰ ویژگی.



شکل ۴-۴: نمودارهای بهترین برازندگی و متوسط برازندگی در ۲۰ تکرار الگوریتم برای کاهش ویژگی‌های حروف بزرگ از ۸۰ به ۳۰ ویژگی.



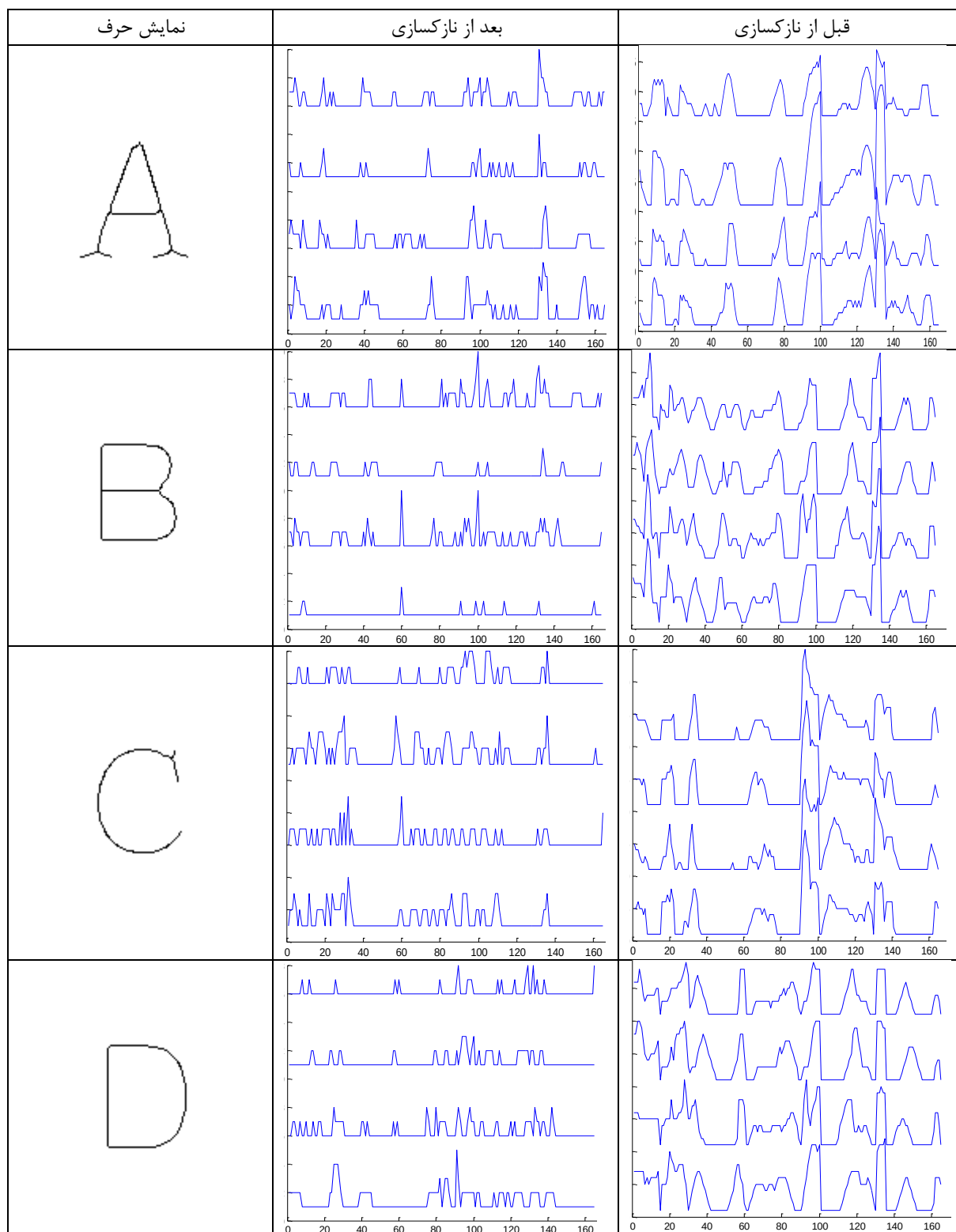
شکل ۴-۵: نمودارهای بهترین برازندگی و متوسط برازندگی در ۲۰ تکرار الگوریتم برای کاهش ویژگی‌های حروف کوچک.

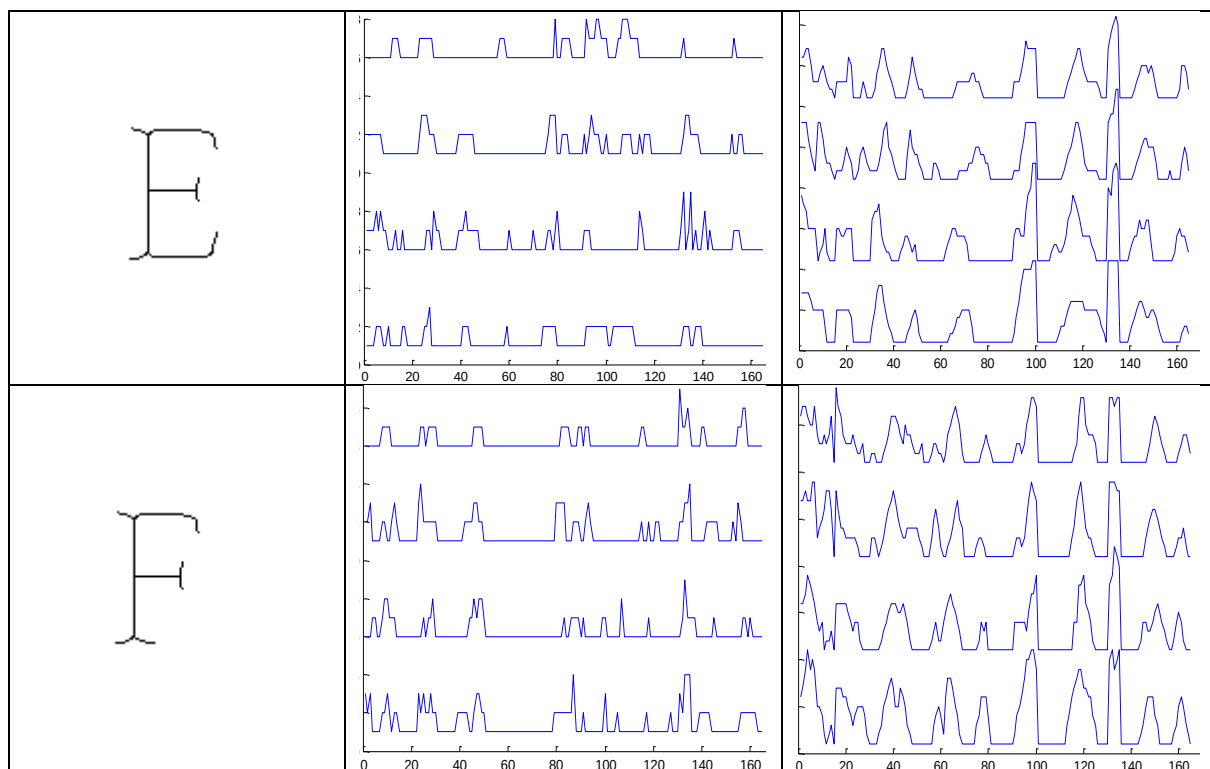
جدول ۴-۸، روش پیشنهادی را با سایر روش‌های موجود مقایسه می‌کند. روش پیشنهادی نسبت به روش‌های دیگر تعداد فونت‌های بیشتری را مورد آزمایش قرار داده است. تعداد ۴۸۶۲ نمونه از داده‌ها دقیقاً در ۱۱ اندازه و ۱۷ فونتی هستند که روش‌های دیگر نیز به کار گرفته‌اند. باقی داده‌ها در ۷ فونت جدید در روش پیشنهادی تست شده‌اند. تعداد کم نمونه‌های آموزشی، سادگی روش کلاس‌بندی، مقاوم بودن نسبت به فونت (با افزایش تعداد فونت‌ها دقت تشخیص به شدت تغییر نکرده است)، مقاوم بودن نسبت به حالات حروف (پررنگ، زاویه‌دار) و تعداد کم ویژگی‌های مورد استفاده از مزایای روش پیشنهادی است.

جدول ۴-۸: مقایسه‌ی روش پیشنهادی تشخیص حروف با سایر روش‌های موجود در این حوزه.

تعداد نمونه‌های آموزشی و آزمایشی	روش کلاس‌بندی	مقاوم بودن نسبت به حالت (پررنگ، زاویه‌دار)	تعداد فونت‌های آزمایش شده	تعداد ویژگی‌ها	دقت
Dhandra, et al.[60] ۳۶۴ آموزش - ۷۲۸۰ تست	درخت تصمیم	×	۷	۵۲	٪۹۹,۶۵
Rani et al.[61] ۲۴۳۱ آموزش - ۴۸۶۲ تست	ماشین بردار پشتیبان	×	۱۷	۱۸۹	٪۹۶,۴۷
Rani et al.[61] ۲۴۳۱ آموزش - ۴۸۶۲ تست	ماشین بردار پشتیبان	×	۱۷	۲۰۰	٪۹۸,۰۸
روش پیشنهادی ۵۲ آموزش - ۴۸۶۲ تست	شبکه عصبی نگاشت خودسازمانده	✓	۱۷	۸۰	٪۹۸,۶۸
روش پیشنهادی ۵۲ آموزش - ۶۱۶۲ تست	شبکه عصبی نگاشت خودسازمانده	✓	۲۴	۸۰	٪۹۸,۵۶
روش پیشنهادی ۵۲ آموزش - ۶۱۶۲ تست	شبکه عصبی نگاشت خودسازمانده	✓	۲۴	۳۰	٪۹۷,۷۳

در برخی از روش‌ها برای ساده شدن محاسبات و استخراج ویژگی‌های کاراکترها، عمل نازک‌سازی را انجام می‌دهند؛ به طوری که موفقیت بسیاری در نتایج به دست آمده نیز حاصل می‌شود [۷۶, ۷۷, ۷۸, ۷۹, ۸۰]. به همین سبب، در سیستم پیشنهادی کم شدن ضخامت ارقام و حروف (انجام عمل نازک‌سازی) را نیز بررسی کردیم. شکل ۴-۶، نمونه‌ای از حروف A تا F نازک شده و نمودار بردار ویژگی آنها را قبل از نازک‌سازی و بعد از آن، نشان می‌دهد. مشاهده می‌شود که بردارهای ویژگی این حروف بعد از نازک‌سازی با وجودی که به دسته‌های مجزا از یکدیگر تعلق دارند، به یکدیگر بسیار شبیه هستند و این عامل سبب می‌شود که سیستم در شناسایی آنها موفق عمل نکند. بنابراین، می‌توان گفت چون در عمل کاراکترها به صورت نازک وجود ندارند، شاید به تست و آزمایش روش پیشنهادی بر روی این گونه نمونه‌ها نیز نیازی نباشد.

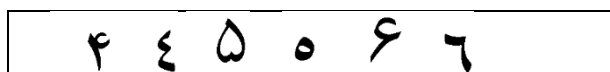




شکل ۴-۶: نمونه‌ای از حروف A تا F به صورت نازک شده و بردارهای ویژگی قبل و بعد از نازک‌سازی آنها.

۴-۲-۲- نتایج حاصل از دسته‌بندی ارقام فارسی

روش پیشنهادی را بر روی ارقام فارسی نیز آزمایش کردیم. پایگاه داده‌ی به‌کار رفته در این آزمایش، حاوی ۲۰۰ نمونه از هریک از ارقام می‌باشد که با فونت‌های B Zar، B Nazanin، B Mitra، B، B Lotus، B Koodak، B Ferdosi، B Fantezy، B Compset، Hamid، B Badr، B Roya، Davat، Shiraz، B Titr، B Traffic، B Yagut، Baran، Karim، B Elham و B Esfehan در اندازه‌های ۱۴ تا ۳۲ (اندازه‌های زوج)، به صورت نرمال و پررنگ ایجاد شده‌اند که در مجموع تعداد نمونه‌های پایگاه داده به ۲۰۰۰ می‌رسد. در پیوست ۲، نمونه فونت‌های مختلف هر یک از ارقام در پایگاه داده نشان داده شده است. برای تهیه‌ی مجموعه نمونه‌های آموزشی، از هر رقم تنها یک نمونه انتخاب شده است؛ اما از آنجایی که در فارسی ارقامی مانند ۶ و ۴ و ۵ به دو صورت مختلف نوشته می‌شوند (شکل ۴-۷)، نیاز است که برای آموزش آنها به سیستم، از هر شکلی یک نمونه داشته باشیم.



شکل ۴-۷: انواع شکل‌های ۴ و ۵ و ۶ در زبان فارسی.

برای تصویر هر رقم با استخراج کوچکترین مستطیل محاط بر رقم مربوطه، خود رقم را از پس‌زمینه جدا می‌کنیم. از آنجایی که فونت و سایز ارقام مختلف یکسان نیستند؛ سایز مستطیل‌های محاطی آن‌ها نیز یکسان نخواهند بود؛ بنابراین، تصاویر به اندازه‌ی 30×30 نرمال می‌شوند. در روش پیشنهادی، به ازای هر رقم، مجموع تعداد یک‌ها در هریک از سطرها و ستون‌ها محاسبه می‌شود. با توجه به آنکه تصویر هر رقم در مرحله پیش پردازش در اندازه 30×30 ذخیره می‌شود، هر رقم به یک سری زمانی با طول ۶۰ تبدیل می‌گردد.

همانند بازشناسی ارقام و حروف انگلیسی، برای بازشناسی ارقام فارسی از شبکه SOM با معیار شباهت Jensen استفاده کردیم. جدول ۴-۹، نتایج حاصل از دسته‌بندی را بر اساس روابط (۴-۲)، (۴-۳) و (۴-۴) نشان می‌دهد. جدول ۴-۱۰، دقت دسته‌بندی هریک از ارقام را نشان می‌دهد. این جدول نشان می‌دهد از مجموع ۲۰۰۰ نمونه‌ی موجود در پایگاه داده، روش پیشنهادی در شناسایی رقم ۲ ضعیف عمل کرده است. مقادیر AC و TFM به دست آمده به ترتیب عبارتند از: ۹۷,۵٪ و ۰,۹۵۱۱.

جدول ۴-۹: دقت روش پیشنهادی بر روی ارقام فارسی.

	عدد نه	عدد هشت	عدد هفت	عدد شش	عدد پنج	عدد چهار	عدد سه	عدد دو	عدد یک	عدد صفر
PR	۱	۱	۰,۹۵	۰,۹۷	۰,۹۹	۰,۹۹	۰,۹۴	۰,۹۶	۰,۹۸	۰,۹۷
R	۰,۹۹	۰,۹۹	۰,۹۶	۱	۱	۰,۹۸	۰,۹۶	۰,۹۰	۰,۹۷	۱
FM	۰,۹۹	۰,۹۹	۰,۹۵	۰,۹۸	۰,۹۹	۰,۹۹	۰,۹۵	۰,۹۲	۰,۹۷	۰,۹۸

جدول ۴-۱۱، مقایسه‌ی روش پیشنهادی را با دو روش دیگر شناسایی ارقام فارسی نشان می‌دهد. همانطور که مشاهده می‌شود تعداد کم نمونه‌های آموزشی، تشخیص نمونه‌های نرمال و پررنگ و مقاوم بودن نسبت به اندازه در کنار تغییر فونت نمونه‌ها از برتری‌های روش پیشنهادی نسبت به سایر روش‌ها است.

جدول ۴-۱۰: نتیجه دسته‌بندی روش پیشنهادی برای شناسایی ارقام فارسی.

ارقام	دقت دسته‌بندی
صفر	٪۱۰۰
یک	٪۹۶,۶
دو	٪۹۰
سه	٪۹۷,۷
چهار	٪۹۸,۸
پنج	٪۱۰۰
شش	٪۱۰۰
هفت	٪۹۶,۱۱
هشت	٪۹۹,۴
نه	٪۹۸,۸
دقت کل	٪۹۷,۵

جدول ۴-۱۱: مقایسه‌ی روش پیشنهادی با سایر روش‌های موجود شناسایی ارقام فارسی.

تعداد نمونه‌های تست و آموزشی	روش کلاس‌بندی	شناسایی حالات مختلف ارقام (نرمال، پررنگ)	دقت دسته‌بندی
H. EbrahimNezhad, GH. A. Montazer, N. Jafari [55] ۲۱۶ تست - ۲۱۶ آموزش	ترکیبی فازی	نرمال	٪۹۷,۵
GH. A. Montazer, H. Qahri Saremi, V. Khatibi [56] ۱۰۵ تست - ۱۲۳ آموزش	نئورو فازی	نرمال	٪۹۵
روش پیشنهادی ۳۰۰۰ تست - ۱۳ آموزش	شبکه نگاشت خودسازمانده	نرمال، پررنگ	٪۹۷,۵

۴-۳- نتیجه‌گیری

در این فصل، نتایج حاصل از اعمال روش پیشنهادی بر روی پایگاه داده‌های ایجاد شده را بیان کردیم. همانطور که گفته شد، ارقام انگلیسی با بردار ویژگی ۹۰ تایی با دقت ٪۹۹,۵۵ و پس از کاهش ویژگی‌ها به ۴۰، با دقت ٪۹۹,۴۴ دسته‌بندی شدند. دقت ٪۹۸,۴۳ با بردار ویژگی ۱۶۵ تایی، دقت

۹۸,۵۶٪ با بردار ویژگی ۸۰ تایی و دقت ۹۷,۷۳٪ با بردار ویژگی ۳۰ تایی پس از دسته‌بندی حروف انگلیسی حاصل شد. همچنین، روش پیشنهادی ارقام فارسی را نیز با دقت ۹۷,۵٪ دسته‌بندی کرد که در ۲۰ نوع فونت متفاوت بودند.

فصل پنجم

نتیجه‌گیری و پیشنهادات

۵-۱- نتیجه گیری

در این پایان نامه، در فصل اول سیستم‌های بازشناسی نوری حروف را معرفی و مراحل مختلف یک سیستم بازشناسی را بیان کردیم. از آنجایی که سیستم بازشناسی ارقام و حروف، یکی از زمینه‌های فعال و به روز تحقیقاتی به شمار می‌آید، در فصل دوم به توضیح و بیان روش‌های موجود برای بازشناسی ارقام و حروف و نقاط ضعف آنها پرداختیم. همانطور که بیان شد، تحقیقات به مراتب کمتری در حوزه شناسایی حروف در چندین فونت و اندازه، انجام شده است. بنابراین، هدف این تحقیق ارائه‌ی روشی است که بتواند ارقام و حروف مختلف را با هر نوع فونت و اندازه شناسایی نماید. روش ارائه شده در این پایان نامه، با بهره‌گیری از یک شبکه‌ی عصبی ساده، بدون نیاز به یک پایگاه داده‌ی آموزشی بزرگ توانست با دقت ۹۹,۴۵٪ ارقام انگلیسی و ۹۷,۷۳٪ حروف انگلیسی را تشخیص دهد. در این روش، در مرحله‌ی پیش‌پردازش، ابتدا تصاویر ارقام و حروف به اندازه‌ی ۴۵ درجه در جهت خلاف عقربه‌های ساعت چرخانده شده و یکسان‌سازی می‌شوند؛ چون برای استخراج ویژگی‌ها، تصویر را از میانه‌ی آن به دو نیمه تقسیم می‌کنیم، چرخاندن تصویر سبب می‌شود اختلاف بین حروف مشابه از دسته‌های مختلف مانند B و D یا ارقام مشابه مانند 0 و 9، زیاد شود و معیار شباهت آنها را مشابه تشخیص نداده و در یک دسته قرار ندهد.

با تقسیم تصاویر به دو قسمت از میانه‌ی تصویر، برای استخراج ویژگی‌های ارقام، با جاروب کردن داده‌ها در جهت سطری و ستونی در هریک از دو نیمه‌ی تصویر، ویژگی‌هایی استخراج می‌شود. این ویژگی‌ها به گونه‌ای هستند که به ازای فونت‌ها و یا اندازه‌های مختلف یک رقم مشابه می‌باشند. این بردار ویژگی به دست آمده برای ارقام به طول ۹۰ است. برای شناسایی حروف، دو مجموعه ویژگی استخراج می‌کنیم. برای تشکیل مجموعه‌ی اول ویژگی‌ها، نیمه‌ی بالایی تصویر را به نسبت $\frac{2}{3}$ و $\frac{1}{3}$ تقسیم کرده و مجموع مقادیر سطرها و ستون‌ها در هریک از این بخش‌ها محاسبه می‌شود. نیمه‌ی پایینی تصویر را هم در جهت افقی و هم در جهت عمودی به دو نیمه تقسیم کرده و با جاروب کردن داده‌ها در جهت‌های سطری و ستونی، مجموعه ویژگی‌های دوم را ایجاد می‌کنیم. این دو مجموعه ویژگی، بردار ویژگی ۱۶۵ تایی را برای حروف می‌سازد؛ بنابراین، برای کاهش ابعاد بردار ویژگی، از الگوریتم وراثتی استفاده می‌کنیم. با استفاده از این الگوریتم، توانستیم بدون کاهش زیاد در میزان دقت، ۳۰ ویژگی از این مجموعه ویژگی‌ها را برای حروف و ۴۰ ویژگی را برای ارقام انتخاب کرده و سرعت شناسایی را افزایش دهیم.

برای دسته‌بندی و تشخیص ارقام و حروف مختلف در این تحقیق، از یک شبکه‌ی عصبی نگاشت خودسازمانده (SOM) استفاده شده است که برای تشخیص کلاس مربوط به یک داده، یک معیار شباهت به کار می‌رود. ایده‌ی استفاده از معیار شباهت، از عملکرد انسان در بازشناسی کاراکترها گرفته شده است. همانطور که اگر از شخصی خواسته شود کاراکتر خاصی را در میان کاراکترهای موجود در فونت‌ها و اندازه‌های مختلف، بازشناسی کند، از شباهت بین کاراکتر موردنظر با کاراکتری که پیش از آن دیده است، استفاده کرده و شخص نیازی ندارد که شکل تمام فونت‌ها را از پیش بداند؛ بلکه با استفاده از شباهت، آن کاراکتر خاص را معرفی می‌کند. بنابراین، در این تحقیق تنها با آموزش یک یا دو نمونه از هریک از کاراکترها به سیستم پیشنهادی، سایر کاراکترها با استفاده از معیار شباهت و شبکه عصبی نگاشت خودسازمانده شناسایی می‌شوند. ارزیابی‌های انجام شده در این تحقیق نشان می‌دهد که روش پیشنهادی پس از آموزش می‌تواند با دقت خوبی ارقام و حروف انگلیسی را در هر اندازه و فونتی شناسایی نماید. بررسی‌های ما نشان می‌دهد که با بهره‌گیری از مجموعه ویژگی بهتر و معیار شباهت مناسب‌تر می‌توان دقت روش پیشنهادی را بهبود بخشید. مقایسه‌ی روش پیشنهادی با سایر روش‌های موجود شناسایی ارقام و حروف، برتری آن را با کم بودن تعداد داده‌های آموزشی، شناسایی نمونه‌های پررنگ^۱ و ایتالیک، مقاوم بودن نسبت به فونت و ساده بودن روش کلاس‌بندی نشان می‌دهد.

۵-۲- پیشنهادات

برای ادامه‌ی کار در زمینه‌ی بازشناسی می‌توان موارد زیر را پیشنهاد داد:

- تلاش برای تقویت هر یک از مراحل بازشناسی متن، مثلاً در مرحله پیش‌پردازش برای باینری کردن، حذف نویز، اصلاح چرخش. به دست آوردن یک تصویر با کیفیت بالا و بدون نویز از مرحله‌ی پیش‌پردازش یا قطعه‌بندی بدون نقص متون، دقت عملکرد سیستم بازشناسی را افزایش می‌دهد؛ بنابراین، با بهبود هریک از این مراحل، می‌توان بهبود عملکرد کل سیستم را در بازشناسی کاراکترها نتیجه گرفت.

¹ Bold

- آزمایش روش پیشنهادی بر روی پایگاه داده‌های بزرگتر شامل انواع فونت‌ها و اندازه‌های بیشتر. هرچه تعداد فونت‌ها و اندازه‌ی نمونه‌های مورد آزمایش بیشتر شود، استفاده از سیستم ارائه شده در کاربردهای مختلف به دلیل کارایی بالاتر پیشنهاد می‌گردد.
- بهبود روش با ایجاد مشخصه‌های دقیق‌تر برای بالابردن دقت بازشناسی.
- کاهش بیشتر ویژگی‌ها بدون تغییر در میزان دقت شناسایی. از میان ویژگی‌های استخراجی، تعداد ۳۰ ویژگی را با استفاده از الگوریتم وراثتی انتخاب کردیم اما ویژگی‌های حاصله ممکن است با ترکیب با یکدیگر، قابل کم شدن باشند و بعد کمتری برای ویژگی‌ها ایجاد کنند.
- آزمایش روش پیشنهادی با سایر معیارهای شباهت.
- تعمیم و استفاده روش پیشنهادی برای شناسایی و دسته بندی سایر زبان‌های موجود در دنیا بدون محدودیت در اندازه و فونت.
- طراحی سیستمی که قابلیت شناسایی تصاویر خاکستری و رنگی را (بدون باینری کردن) داشته باشد.
- طراحی سیستمی که بتواند تصاویر با رزولوشن پایین را نیز با استفاده از روش پیشنهادی به خوبی دسته‌بندی کند.
- استفاده از ایده‌ی پیشنهادی برای جداسازی کاراکترهای دست‌نوشته در زبان‌های مختلف. می‌توان مفهوم شباهت را مانند استفاده در تشخیص کاراکترهای چاپی، برای شناسایی کاراکترهای دست‌نوشته به کار برد اما نیاز است که معیار شباهت متناسب برای کاربرد موردنظر را بیابیم.
- استفاده از شبکه‌های ترکیبی به منظور افزایش دقت بازشناسی. می‌توان از الگوریتم‌های آدابوست^۱ و ترکیب سایر کلاسه‌بندها با شبکه عصبی نگاشت خودسازماندهی معرفی شده استفاده کرد تا نمونه‌هایی که اشتباه دسته‌بندی شده‌اند، به طور صحیح شناسایی شده و به دقت بالاتری رسید.
- ارائه‌ی روشی برای انتخاب نماینده‌ی مناسب برای هر یک از ارقام و حروف جهت آموزش. اگر بتوان نماینده‌ی منحصر به فرد و متناسب با هریک از ارقام و حروف را به درستی انتخاب کرد، می‌توان دقت بازشناسی در روش پیشنهادی را بهبود داد.

¹ Adaboost

پیوست ۱

پ-۱-۱- معیارهای شباهت

به منظور سنجش میزان شباهت بین دو سری زمانی می‌توان آنها را به یک معیار شباهت اعمال نمود. معیار شباهت‌های متفاوتی وجود دارند که هر کدام برای کاربردهای خاصی مناسب‌تر هستند. معیارهای شباهت به سه دسته تقسیم می‌شوند:

۱. معیارهای فاصله (Distance measures)

۲. معیارهای نزدیکی (Proximity measures)

۳. معیارهای دودویی (Binary measures)

پ-۱-۱-۱- معیارهای فاصله

این معیارها، نمونه‌های دو سری زمانی را نظیر به نظیر (one-by-one) با یکدیگر مقایسه می‌کنند. در این گروه، معیار شباهت $d(x,y)$ دو دنباله‌ی x و y ، باید شروط زیر را ارضا کند:

✓ مثبت بودن: برای هر x و y ، $d(x,y)$ یک عدد حقیقی بزرگتر مساوی صفر است. اگر $d=0$ اگر و تنها اگر $x=y$.

✓ تقارن: $d(x,y)=d(y,x)$.

✓ نامساوی مثلثی: $d(x,z) \leq d(x,y) + d(y,z)$.

۱. فاصله اقلیدسی

فاصله‌ی بین دو دنباله، برابر مجذور جمع اختلافات مربع مقادیر دو دنباله است:

$$Euclid(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (1-1)$$

۲. معیار Chebychev

فاصله‌ی بین دو دنباله، ماکسیمم قدرمطلق اختلاف مقادیر دنباله‌هاست:

$$Chebychev(x,y) = \max |x_i - y_i| \quad (2-1)$$

۳. فاصله همینگ

فاصله‌ی دو دنباله، جمع اختلافات مقادیر دو دنباله است:

$$Block(x,y) = \sum_i |x_i - y_i| \quad (3-1)$$

۴. معیار Minkowski

فاصله‌ی بین دو دنباله، p امین ریشه‌ی جمع قدرمطلق اختلافات به توان p مقادیر دو دنباله است. فاصله Minkowski می‌تواند برخی معیارهای شباهت این گروه را نشان دهد؛ برای مثال با قراردادن p برابر ۱ و ۲، به ترتیب فاصله همینگ و فاصله اقلیدسی به دست می‌آید.

$$Minkowski(x, y) = \left(\sum_i |x_i - y_i|^p \right)^{1/p} \quad (4 - 1)$$

پ-۱-۱-۲- معیارهای نزدیکی

معیارهای این دسته، ویژگی‌های سراسری دو سری زمانی را در نظر می‌گیرند و برای اغلب کاربردهای خاص قابل استفاده هستند. همچنین، این معیارها نتایج ذاتی بیشتری در مقایسه با معیارهای گروه قبل دارند.

۱. معیار Cosine

این معیار، یک معیار شباهت مبتنی بر زاویه است که شباهت بین نمونه‌های نرمال شده‌ی دو دنباله را شرح می‌دهد. به عبارت دیگر، تنها جهت‌های بردارهای واحد که دو دنباله را نشان می‌دهند در این معیار در نظر گرفته می‌شوند:

$$Cosine(x, y) = \frac{\sum_i x_i y_i}{\sqrt{(\sum_i x_i^2)(\sum_i y_i^2)}} \quad (5 - 1)$$

۲. معیار Correlation

این معیار، هم‌بستگی بین دو دنباله x و y را نشان می‌دهد:

$$Correlation(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6 - 1)$$

$$\bar{x} = \frac{1}{n} \sum_i x_i \quad \bar{y} = \frac{1}{n} \sum_i y_i$$

۳. معیار Chi-square

این معیار، به فرکانس‌های کل دو دنباله بستگی دارد. مقادیر موردانتظار، مدل استقلال دنباله‌های x و y هستند.

$$Chisq(x, y) = \sqrt{\sum_i \frac{(x_i - E(x_i))^2}{E(x_i)} + \sum_i \frac{(y_i - E(y_i))^2}{E(y_i)}} \quad (7-1)$$

۴. معیار Jensen

خروجی این معیار مقداری بین ۰ و ۱ است که اگر دو سری کاملاً یکسان باشند، مقدار ۰ برگردانده می‌شود. این تابع برای مقایسه‌ی دنباله‌هایی با مقادیر مثبت استفاده می‌شود.

$$\frac{1}{2} \sum_{i=1}^k \left\{ p'_i \log_2 p'_i + q'_i \log_2 q'_i - (p'_i + q'_i) \log_2 \left(\frac{p'_i + q'_i}{2} \right) \right\} \quad (8-1)$$

$$\text{where } p'_i = p_i / \sum_{i=1}^k p_i, \quad q'_i = q_i / \sum_{i=1}^k q_i,$$

پ-۱-۱-۳- معیارهای دودویی

در برخی مسائل، نیاز به مقایسه‌ی دنباله‌هایی با مقادیر دودویی است. معیارهای شباهت معرفی شده در دو گروه قبل ممکن است برای مقایسه‌ی دنباله‌های دودویی مناسب نباشند. برای کمیت‌گذاری شباهت بین این دنباله‌ها چهار متغیر مجموع a,b,c,d بدین صورت استفاده می‌شوند: تطابق مثبت (۱-۱)، عدم تطابق چپ (۰-۱)، عدم تطابق راست (۱-۰) و تعداد تطابقات منفی (۰-۰). معیارهای دودویی با تغییر متغیرهای مجموع، قابل توسعه و استفاده برای داده‌های غیر دودویی نیز هستند.

۱. معیار Jaccard

یکی از محبوب‌ترین معیارها است که به نام نرخ شباهت نیز شناخته می‌شود:

$$Jaccard(x, y) = \frac{a}{a + b + c} \quad (9-1)$$

۲. معیار شباهت Ochiai

این معیار شکل دودویی معیار Cosine است. مقدار این معیار بین ۰ و ۱ تغییر می‌کند:

$$Ochiai(x, y) = \sqrt{\left(\frac{a}{a+b} \right) \left(\frac{a}{a+c} \right)} \quad (10-1)$$

۳. معیار Hamman

مقدار این معیار بین -۱ و +۱ تغییر می‌کند و به صورت زیر تعریف می‌شود:

$$Hamman(x, y) = \frac{(a + d) - (b + c)}{a + b + c + d} \quad (11 - 1)$$

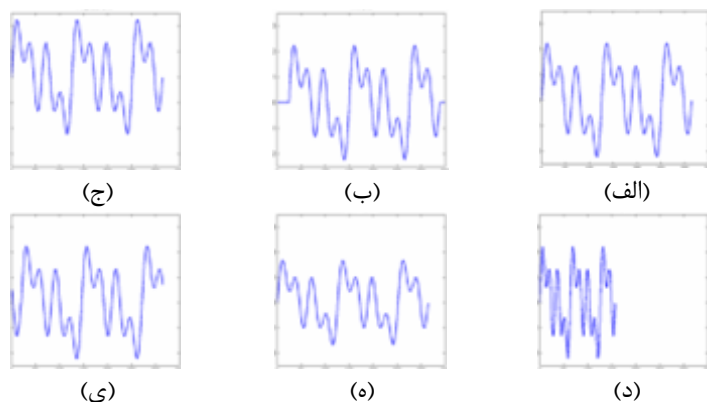
۴. معیار Goodman and Kruskal Lambda

این معیار، کاهش نسبی در خطا را با استفاده از یک دنباله برای پیش‌بینی دنباله‌ی دیگر، کمیت‌گذاری می‌کند؛ زمانی که جهت‌های پیش‌بینی، اهمیت یکسان دارند. Lambda، بین ۰ و ۱ تغییر می‌کند.

$$\begin{aligned} t_1 &= \max(a, b) + \max(c, d) + \max(b, d) \\ t_2 &= \max(a + c, b + d) + \max(a + b, c + d) \\ Lambda(x, y) &= \frac{t_1 - t_2}{2(a + b + c + d) - t_2} \end{aligned} \quad (12 - 1)$$

پ-۱-۲- مقایسه‌ی معیارهای شباهت

همانطور که گفته شد، معیارهای نزدیکی ویژگی‌های سراسری دو سری را در نظر می‌گیرند؛ بنابراین در مقایسه‌ی دو سری لازم است بررسی گردد که سری دوم نسبت به سری اول چه تغییری کرده است. این تغییرات شامل شیف در زمان، شیف در دامنه، مقیاس^۱ زمان، مقیاس دامنه و تغییر فاز باشد.



شکل پ-۱: (الف) سیگنال اصلی، (ب) شیف در زمان، (ج) شیف دامنه، (د) مقیاس زمان، (ه) مقیاس دامنه، (ی) تاخیر فاز.

هر معیار شباهتی نسبت به برخی تغییرات مقاوم است؛ یعنی اگر تغییرات گفته شده به سری اولیه اعمال گردد، سری حاصل را مشابه آن می‌پندارد. جدول (۱-۱) معیارهای شباهت و مقاوم بودن نسبت به تغییرات را نشان می‌دهد.

¹ scale

معیارها	ویژگی‌ها				
	تأخیر فاز	مقیاس دامنه	مقیاس زمان	ثبوت دامنه	ثبوت در زمان
معیارهای فاصله	x	x	x	x	x
Chi-square	✓	x	x	x	x
جنسن	x	✓	x	x	x
Cosine	x	✓	x	x	x
Correlation	x	x	x	✓	x

جدول پ-۱: مقایسه ویژگی‌های معیارهای مختلف شباهت.

پیوست ۲

در این پیوست، نمونه فونت‌های استفاده شده در پایگاه داده نشان داده شده‌اند.

پ-۲-۱- نمونه‌هایی از فونت‌های مختلف ارقام موجود در پایگاه داده

Font	Numerals	Font	Numerals
Times	0123456789	Calibri	0123456789
Arial	0123456789	Times new roman	0123456789
Palatino	0123456789	Georgia	0123456789
Cambria	0123456789	ArialRoundedMT Bold	0123456789
Lucida Sans	0123456789	AvantGarde	0123456789
Comic	0123456789	Century School Book	0123456789
Garamond	0123456789	Verdana	0123456789
Dejavu Sans	0123456789	Century Gothic	0123456789
Coolvetica	0123456789	CourierPS	0123456789
Microsoft Sans Serif	0123456789	BookMan	0123456789
New Century School Book	0123456789	Courier	0123456789
Shruti	0123456789	Trebuchet MS	0123456789
Estrangelo Edessa	0123456789	Lucida Bright	0123456789
Lucida Console	0123456789	Book Antiqua	0123456789
Tahoma	0123456789	Sylfean	0123456789

پ-۲-۲: نمونه‌هایی از فونت‌های مختلف حروف بزرگ موجود در پایگاه داده

حروف	نوع فونت	حروف	نوع فونت
ABCDEFGHIJKLM NOPQRSTUVWXYZ Z	Times	ABCDEFGHIJKLMNO PQRSTUVWXYZ	Calibri
ABCDEFGHIJKLM NOPQRSTUVWXYZ Z	Times new roman	ABCDEFGHIJKLMNO PQRSTUVWXYZ	CourierPS
ABCDEFGHIJKLM NOPQRSTUVWXYZ Z	Comic	ABCDEFGHIJKLM NOPQRSTUVWXYZ	ArialRounded MTBold
ABCDEFGHIJKLMN OPQRSTUVWXYZ	Georgia	ABCDEFGHIJKLM NOPQRSTUVWXYZ	Bodoni MT
ABCDEFGHIJKLMN OPQRSTUVWXYZ	Lucida Sans Typewriter	ABCDEFGHIJKLMN OPQRSTUVWXYZ	Rockwell
ABCDEFGHIJKLMNO PQRSTUVWXYZ	Cambria	ABCDEFGHIJKLM OPQRSTUVWXYZ	CenturyScho olBook
ABCDEFGHIJKLMN OPQRSTUVWXYZ	Harry P	ABCDEFGHIJKL MNOPQRSTUVW XYZ	Copperplate Gothic Light
ABCDEFGHIJKLM NOPQRSTUVWXYZ Z	Garamond	ABCDEFGHIJKLMN OPQRSTUVWXYZ	Verdana
ABCDEFGHIJKLM NOPQRSTUVWXYZ Z	Arial	ABCDEFGHIJKLMNO PQRSTUVWXYZ	Coolvetica
ABCDEFGHIJKLM NOPQRSTUVWXYZ Z	Palatino	ABCDEFGHIJKLM NOPQRSTUVWXYZ	BookMan
ABCDEFGHIJKLM NOPQRSTUVWXYZ Z	MicrosoftSansSer if	ABCDEFGHIJKLMNO PQRSTUVWXYZ	Tahoma
ABCDEFGHIJKLM NOPQRSTUVWXYZ Z	NewCenturyScho olBook	ABCDEFGHIJKLMN OPQRSTUVWXYZ	Felix Titling

پ-۲-۲- نمونه‌هایی از فونت‌های مختلف حروف کوچک موجود در پایگاه داده

حروف	نوع فونت	حروف	نوع فونت
abcdefghijklm nopqrstuvwxyz	Lucida Sans Typewriter	abcdefghijklmnop qrstuvwxyz	Rockwell
abcdefghijklmnopqr stuvwxyz	MicrosoftSansSerif	abcdefghijklmnopqr stuvwxyz	Tahoma
abcdefghijklm nopqrstuvwxyz	Harry P	abcdefghijklmnopqrstuvwxy z	Haettenschweil er
abcdefghijklmnopqr stuvwxyz	Times	abcdefghijklmnopqrs tuvwxyz	Calibri
abcdefghijklmnopqr stuvwxyz	Times new roman	abcdefghijklmn opqrstuvwxyz	CourierPS
abcdefghijklmno pqrstuvwxyz	Comic	abcdefghijklmnop qrstuvwxyz	ArialRounded MTBold
abcdefghijklmnop qrstuvwxyz	Georgia	abcdefghijklmnopqrs tuvwxyz	Bodoni MT
abcdefghijklmnopqr stuvwxyz	NewCenturySchool Book	abcdefghijklmnopqrs tuvwxyz	Tempus Sans ICT
abcdefghijklmnop qrstuvwxyz	Cambria	abcdefghijklmnopq rstuvwxyz	CenturySchool Book
abcdefghijklmopqrst uvwxyz	Garamond	abcdefghijklmno pqrstuvwxyz	Verdana
abcdefghijklmnop qrstuvwxyz	Arial	abcdefghijklmnopqrstuv wxyz	Coolvetica
abcdefghijklmnop qrstuvwxyz	Palatino	abcdefghijklmnop qrstuvwxyz	BookMan

پ-۲-۳: نمونه‌هایی از فونت‌های مختلف حروف بزرگ پررنگ و زاویه‌دار موجود در پایگاه داده.

حروف	نوع فونت
<i>ABCDEFGHIJKLMNOPQRSTUVWXYZ</i>	Times New Roman
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>ABCDEFGHIJKLMNOPQRSTUVWXYZ</i>	Bodoni MT
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>ABCDEFGHIJKLMNOPQRSTUVWXYZ</i>	Calibri
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>ABCDEFGHIJKLMNOPQRSTUVWXYZ</i>	Cambria
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>ABCDEFGHIJKLMNOPQRSTUVWXYZ</i>	Century Gothic
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>ABCDEFGHIJKLMNOPQRSTUVWXYZ</i>	Rockwell
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>ABCDEFGHIJKLMNOPQRSTUVWXYZ</i>	Arial
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>ABCDEFGHIJKLMNOPQRSTUVWXYZ</i>	Tahoma
ABCDEFGHIJKLMNOPQRSTUVWXYZ	

پ-۲-۴: نمونه‌هایی از فونت‌های مختلف حروف کوچک پررنگ و زاویه‌دار موجود در پایگاه

داده

حروف	نوع فونت
<i>abcdefghijklmnopqrstuvwxyz</i>	Times New Roman
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>abcdefghijklmnopqrstuvwxyz</i>	Bodoni MT
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>abcdefghijklmnopqrstuvwxyz</i>	Calibri
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>abcdefghijklmnopqrstuvwxyz</i>	Cambria
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>abcdefghijklmnopqrstuvwxyz</i>	Century Gothic
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>abcdefghijklmnopqrstuvwxyz</i>	Rockwell
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
<i>abcdefghijklmnopqrstuvwxyz</i>	Arial
ABCDEFGHIJKLMNOPQRSTUVWXYZ	
ABCDEFGHIJKLMNOPQRSTUVWXYZ	Tahoma

پ-۲-۵: نمونه‌هایی از فونت‌های مختلف ارقام فارسی موجود در پایگاه داده

نوع فونت	اعداد	نوع فونت	اعداد
B Nazanin	۰۱۲۳۴۵۶۷۸۹	B Koodak	۰۱۲۳۴۵۶۷۸۹
B Mitra	۰۱۲۳۴۵۶۷۸۹	B Lotus	۰۱۲۳۴۵۶۷۸۹
B Zar	۰۱۲۳۴۵۶۷۸۹	B Shiraz	۰۱۲۳۴۵۶۷۸۹
B Davat	۰۱۲۳۴۵۶۷۸۹	B Titr	۰۱۲۳۴۵۶۷۸۹
B Roya	۰۱۲۳۴۵۶۷۸۹	B Traffic	۰۱۲۳۴۵۶۷۸۹
B Badr	۰۱۲۳۴۵۶۷۸۹	B Yagut	۰۱۲۳۴۵۶۷۸۹
2 Hamid	۰۱۲۳۴۵۶۷۸۹	2 Baran	۰۱۲۳۴۵۶۷۸۹
B Compset	۰۱۲۳۴۵۶۷۸۹	2 Karim	۰۱۲۳۴۵۶۷۸۹
B Fantezy	۰۱۲۳۴۵۶۷۸۹	B Elham	۰۱۲۳۴۵۶۷۸۹
B Ferdosi	۰۱۲۳۴۵۶۷۸۹	B Esfehan	۰۱۲۳۴۵۶۷۸۹

فهرست مراجع

1. Trier, O.D., Jain, A.K., and Taxt, T., *Feature extraction methods for character recognition - A survey*. Pattern Recognition, 1996. **29**: p. 641-662.
2. Mithe, R., S. Indalkar, Divekar, N., *Optical Character Recognition*. International Journal of Recent Technology and Engineering (IJRTE), March 2013. **2**(1).
3. Nazif, A. and Yamin-Vural Fatos, T., *An overview of character recognition based focused on off-line handwriting*, in *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*. may 2001.
4. Mantas, J., *An overview of character recognition methodologies*. Pattern Recognition, 1986. **19**: p. 425-30.
5. Mori, S., Suen, C.Y., Yamamoto, K., *Historical review of OCR research and development*. in *IEEE*. 1992. **80**:p. 1029-58.
6. Suen, C.Y., Tappert, C.C. and Wakahara, T., *The state of the art in online handwriting recognition*, in *IEEE Trans. Pattern Anal. Machine Intell*. 1990. p. 787-808.
7. Sattar, S.a., *A Technique for the design and implementation of an OCR for printed nastalique text*, in *Department of Computer Science and Information Technology*. PHD Thesis, July 2009, N.E.D. University of Engineering & Technology: Karachi, Pakistan.
۸. باقری، ز.، بازشناسی متن تایپی نوشته شده با قلم *Iranian Sans*، پایان نامه کارشناسی ارشد، دانشگاه شاهرود، دانشکده مهندسی برق و رباتیک، ۱۳۹۲.
۹. فرامرزی، ا.، بازشناسی نوری حروف: مروری بر مباحث نظری وملاحظات کاربردی با تأکید بر مسائل خاص زبان فارسی. ۱۳۸۴، مجلد ۳، صفحات ۳۳-۶۱.
10. Alon, J., Athitsos, V., and Sclaroff, S., *Online and offline character recognition using alignment to prototypes*. in *Eighth International Conference on Document Analysis and Recognition*, 2005.
11. Mohiuddin, K.M. and Mao, J., *Optical Character Recognition*. Wiley Encyclopedia of Electrical and Electronics Engineering, 1999, p.226-236.
12. Alginahi, Y., *Preprocessing Techniques in Character Recognition*, in *Computer and Information Science*, 2010.
13. Gonzalez, R.C., *Digital Image Processing*. 2 ed. USA: Prentice Hall. 2001.
14. O'Gorman, et al., *Document Image Analysis*, in *IEEE Computer Society Press* , Los Alamitos, 1995.
15. Hallale, S.B. and Salunke, G.D., *Twelve Directional Feature Extraction for Handwritten English Character Recognition*, International Journal of Recent Technology and Engineering (IJRTE), 2013. **2**(2).
16. Hull, J. J. and Taylor, S.L., *Document Image Skew Detection Survey and Annotated Bibliography*, World Scientific, 1998, p.40-64.
17. Jain, L.C. and Lazzerini., B., *Knowledge-Based Intelligent Techniques in Character Recognition* , London: CRC Press, 1999.
18. Rehman, A., Dzulkigli, M. and Sulong, G., *Implicit Vs Explicit based Script Segmentation and Recognition: A Performance Comparison on Benchmark Database*. International Journal of Open Problems Compt. Math., 2009. **2**(3).

19. Choudhary, A., *A Review of Various Character Segmentation Techniques for Cursive Handwritten Words Recognition*. International Journal of Information & Computation Technology, 2014. **4**(6): p. 559-564.
20. Cheriet, M., Kharma, N., Liu, C. and Suen, C.Y., *Character Recognition Systems: A Guide for Students and Practitioners*. John Wiley & Sons, 2007.
21. Dongre, V.J. and Mankar, V.H., *A review of Research on Devnagari Character Recognition*. International Journal of Computer Applications, 2010. **12**(2): p. 8-15.
22. Hoffmann, H. and Chamie, M., *Standard Statistical Classification: Basic Principles*, In *The Third Meeting of the Expert Group on International Economic and Social Classifications*. 1999: New York.
23. Yalniz, I.Z. and Manmatha, R., *A Fast Alignment Scheme for Automatic OCR Evaluation of Books*. in *ICDAR'11 Proceedings of the 2011 International Conference on Document Analysis and Recognition*. 2011.
24. Impedovo, D. and Pirlo, G., *Zoning methods for handwritten character recognition: A survey*. Pattern Recognition, 2014. **47**(3): p. 969-981.
25. Bokser, M., *Omnidocument technologies*. Proceedings of the IEEE, 1992. **80**(7): p. 1066-1078.
26. Singh, S. and Hewitt, M., *Cursive digit and character recognition in CEDAR database*. in *15th International Conference on Pattern Recognition*, 2000.
27. Gagné, C. and Parizeau, M., *Genetic engineering of hierarchical fuzzy regional representations for handwritten character recognition*. International Journal of Document Analysis and Recognition (IJДАР), 2006. **8**(4): p. 223-231.
28. Arnold, R. and Miklos, P., *Character recognition using neural networks*. in *11th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2010.
29. Arora, S., Bhattacharjee, D., Nasipuri, M., Malik, L., Kundu, M. and Basu, D.K., *Performance comparison of SVM and ANN for Handwritten Devnagari Character Recognition*. International journal of computer science issues (IJCSI), May 2010. **7**(3).
30. ElKessab, B., Daoui, C., Bouikhanlene, B., Fakir M. and Moro, K., *Extraction Method of Handwritten Digit Recognition Tested on the MNIST Database*. International Journal of Advanced Science and Technology, 2013. **50**.
31. Shrivastava, S. and Singh, M.P., *Performance evaluation of feed-forward neural network with soft computing techniques for hand written English alphabets*. Applied Soft Computing, 2011. **11**(1): p. 1156-1182.
32. Zamora-Martínez, F., Frinken, V., España-Boquera, S., Castro-Bleda, M. J., Fischer, A. and Bunke, H., *Neural network language models for off-line handwriting recognition*. Pattern Recognition, 2014. **47**(4): p. 1642-1652.
33. Kumar, S. and Singh, M.P., *Study of Hopfield neural network with sub-optimal and random GA for pattern recalling of English characters*. Applied Soft Computing, 2012. **12**(8): p. 2593-2600.
34. Namane, A., Maamoun, M., Soubari, E. H. and Meyrueis, P., *CSM-based feature extraction for degraded machine printed character recognition*. in *IEEE 9th International Conference on Cybernetic Intelligent Systems (CIS)*, 2010.

35. Sukhija, S., Panwar, S., and Nain, N., *CRAMM: Character recognition aided by mathematical morphology*. in *IEEE Second International Conference on Image Information Processing (ICIIP)*, 2013.
36. Ben Moussa, S., Zahour, A., Benabdelhafid, A. and Alimi, A. M., *Fractal-based system for Arabic/Latin, printed/handwritten script identification*. in *19th International Conference on Pattern Recognition(ICPR)*, 2008.
37. Nawaz, S.N., Sarfraz, M., Zidouri, A. and Al-Khatib, W. G., *An approach to offline Arabic character recognition using neural networks*. in *Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2003*.
38. Khorsheed, M.S. and Al-Omari, H., *Recognizing cursive Arabic text: Using statistical features and interconnected mono-HMMs*. in *4th International Congress on Image and Signal Processing (CISP)*, 2011.
39. Slimane, F., Kanoun, S., Abed, H. E., Alimi, A. M., Ingold, R. and Hennebert, J., *ICDAR 2011 - Arabic Recognition Competition: Multi-font Multi-size Digitally Represented Text*. in *International Conference on Document Analysis and Recognition (ICDAR)*, 2011.
40. Broumandnia, A. and Shanbehzadeh, J., *Fast Zernike wavelet moments for Farsi character recognition*. *Image and Vision Computing*, 2007. **25**(5): p. 717-726.
41. Angadi, S.A., Kodabagi, M.M., and Jerabandi, M.V., *Character recognition of Kannada text in low resolution display board images using zone wise statistical features*. in *World Congress on Information and Communication Technologies (WICT)*, 2012.
42. Mukarambi, G., Dhandra, B.V. and Hangarge, M., *A Zone Based Character Recognition Engine for Kannada and English Scripts*. *Procedia Engineering*, 2012. **38**(0): p. 3292-3299.
43. Horiuchi, T. and Kato, S., *A Study on Japanese Historical Character Recognition Using Modular Neural Networks*. in *Fourth International Conference on Innovative Computing, Information and Control (ICICIC)*, 2009.
44. Al-Omari, F.A. and Al-Jarrah, O., *Handwritten Indian numerals recognition system using probabilistic neural networks*. *Advanced Engineering Informatics*, 2004. **18**(1): p. 9-16.
45. Clocksin, W.F. *Towards automatic transcription of Syriac handwriting*. in *12th International Conference on Image Analysis and Processing*, 2003.
46. Yanhua, M. and Chuanjun, L., *A Recognition Algorithm for Chinese Character Based on Minimum Distance Classifier*. in *Second International Workshop on Computer Science and Engineering, WCSE '09*, 2009.
47. Zramdini, A. and Ingold, R., *Optical font recognition using typographical features*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998. **20**(8): p. 877-882.
48. راشدی، ع، نظام آبادی پور ح و سریزدی س، *بازشناسی قلمهای فارسی به کمک ضرایب خودهمبستگی*، چهارمین کنفرانس ماشین بینایی و پردازش تصویر ایران. ۱۳۸۵، دانشگاه فردوسی مشهد.
49. Min-Chul, J., Shin, Y.C. and Srihari, S.N., *Multifont classification using typographical attributes*. in *Proceedings of the Fifth International Conference on Document Analysis and Recognition(ICDAR '99)*, 1999.
50. Yong, Z., Tieniu, T. and Yunhong, W., *Font recognition based on global texture analysis*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001. **23**(10): p. 1192-1200.

۵۱. نظام آبادی پور، ح.، سریزدی س. و ا. ابراهیمی، بازشناسی قلمهای فارسی بر پایه فیلترهای گابور، نهمین کنفرانس انجمن کامپیوتر ایران. ۱۳۸۲، صفحات ۳۷۱-۳۷۸.

52. Khosravi, H. and Kabir, E., *Farsi font recognition based on Sobel–Roberts features*. Pattern Recognition Letters, 2010. **31**(1): p. 75-82.
53. Avilés-Cruz, C., Rangel-Kuoppa, R., Reyes-Ayala, M., Andrade-Gonzalez, A. and Escarela-Perez, R., *High-order statistical texture analysis—font recognition applied*. Pattern Recognition Letters, 2005. **26**(2): p. 135-145.
54. Mallikarjun, H., Shashikala, P., Dhandra, B.V., *Multi-font/size Kannada Vowels and Numerals Recognition Based on Modified Invariant Moments*. International Journal of Computer Applications, 2010, p. 126–130.
55. Boveiri, R.H., *Transformation-Invariant Classification of Persian Printed Digits*. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2011. **4**(3).
56. Ebrahimnezhad, H., Montazer, G.A. and Jafari N., *Recognition of Persian Numeral Fonts by Combining the Entropy minimized fuzzifier and fuzzy Grammar*, in *WSEAS Transactions on Artificial intelligence, knowledge Engineering and data bases*. 2007.
57. Montazer, G.A., Saremi, H.Q. and Khatibi, V., *A neuro-fuzzy inference engine for Farsi numeral characters recognition*. Expert Systems with Applications, 2010. **37**(9): p. 6327-6337.
58. Hase, H., Shinokawa, T., Tokai, S. and Suen, C. Y., *A robust method of recognizing multi-font rotated characters*. in *Proceedings of the 17th International Conference on Pattern Recognition, (ICPR) 2004*.
59. Hase, H., Tanabe, K., Thi Hong Ha, T. and Tokai, S., *Multi-Font Rotated Character Recognition Using Periodicity*. in the *Eighth IAPR International Workshop on Document Analysis Systems (DAS '08)*, 2008.
60. Dhandra, B.V., Malemath, V. S., Mallikarjun, H. and Hegadi, R., *Multi-font Numeral Recognition without Thinning based on Directional Density of Pixels*. in *1st International Conference on Digital Information Management*, 2007.
61. Dhandra, B.V., Malemath, V.S., Mallikarjun, H. and Hegadi, R., *Multi-font English Character Recognition based on Modified Invariant Moments*. Journal of Combinatorial Mathematics and Combinatorial Computing, 2008. **76**: p. 153-162.
62. Rani, R., Dhir, R., and Lehal, G.S., *Script Identification of Pre-segmented Multi-font Characters and Digits*. in *12th International Conference on Document Analysis and Recognition (ICDAR)*, 2013.
63. Grafmüller, M. and Beyerer, J., *Performance improvement of character recognition in industrial applications using prior knowledge for more reliable segmentation*. Expert Systems with Applications, 2013. **40**(17): p. 6955-6963.
64. Pratim Roy, P., Pal, U., Lladós, J. and Delalandre, M., *Multi-oriented touching text character segmentation in graphical documents using dynamic programming*. Pattern Recognition, 2012. **45**(5): p. 1972-1983.
65. Lacerda, E.B. and Mello, C.A.B., *Segmentation of connected handwritten digits using Self-Organizing Maps*. Expert Systems with Applications, 2013. **40**(15): p. 5867-5877.

66. Maitre, G., *Experiments with robust similarity measures for OCR*. IDIAP TR June 1995: p. 95-103.
67. Kahya, E., *A new unidimensional search method for optimization: Linear interpolation method*. Applied Mathematics and Computation, 2005. **171**(2): p. 912-926.
68. Sablonnière, P., *Interpolation by quadratic splines on triangles and squares*. Computers in Industry, 1982. **3**(1-2): p. 45-52.
69. Q. Duan, K.D., Price, W.G. and Twizell, E.H., *Weighted rational cubic spline interpolation and its application*. Journal of Computational and Applied Mathematics, 2000. **117**(2): p. 121-135.
70. Dunlop, G.R., *A rapid computational method for improvements to nearest neighbor interpolation*. Computers & Mathematics with Applications 1980. **6**(3): p. 349-353.
71. Hassanpour, H., Darvishi, A. and Khalili, A., *A regression-based approach for measuring similarity in discrete signals*. International Journal of Electronics, 2011. **98**.
72. Tokunaga, K. and Furukawa, T., *Modular network SOM*. Neural Networks, 2009. **22**(1): p. 82-90.
73. D.E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*. 1989: Addison Wesley.
74. Siedlecki, W. and Sklansky J., *A note on genetic algorithm for large scale feature selection*. pattern Recognition Letters, 1989. **10**: p. 335-347.
۷۵. کبیر ا. و نظام آبادی ح., ترکیب ویژگیهای سطح پایین برای طبقه بندی معنایی تصاویر، نشریه علمی پژوهشی انجمن کامپیوتر ایران، ۱۳۸۳. مجلد ۲، صفحات ۳۷-۴۶.
76. Trier, O. D., Jain, A. K. and Taxt, T., *Feature extraction methods for character recognition*, Pattern Recognition, 1996, 29(4): p. 641-662.
77. Chacko, B. P., Vimal Krishnan, V. R., Raju, G. and Babu Anto, P., *Handwritten character recognition using wavelet energy and extreme learning machine*, International Journal of Machine Learning and Cybernetics, June 2012. 3(2): p. 149-161.
78. Ghosh, S. and Bag, S., *A modified thinning strategy to handle junction point distortion for bangla characters*, In IEEE Students' Technology Symposium (TechSym), 2014. p. 52-56.
79. Yi, C. and Tian, Y., *Scene text recognition in mobile applications by character descriptor and structur confirgution*, In IEEE Transactions on Image Processing, 2014, 23(7): p. 2972-2982.
80. Dwivedi, N., Srivastava, K. and Arya, N., *Sanskrit word recognition using Prewitt's operator and support vector classification*, In International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 2013. p. 265-269.

Abstract

OCR plays an important role in different areas such as license plate recognition, security (passport authentication), barcode and bank systems (check). There are many methods and systems for recognizing characters. The purpose of this study proposed a method to recognize letters and digits in a language free limitation in font type and size.

The most existing recognition systems have many training samples and as a result, their computation complexity is high; because they use a set of training samples per font. Also, if these systems have been extended to identify more fonts, the recognition error rate is increased. So, it needs a system with small training samples which can be resistant to different the fonts and Fault diagnosis not much has changed by adding new fonts.

In this thesis, we recognize typed digits and letters in English. Typed characters are written in different fonts and sizes that make similar characters being different with each other; therefore, a system can recognize similar characters with extracting appropriate features form the characters. In this study, we could identify English digits in 30 different font and 11 different sizes with 99.47% accuracy and English letters in 24 different font and 11 different sizes with 97.73% accuracy using a simple feature vector- the sum of values of pixels in each row and column- and a SOM neural network without complex architecture.

Self-Organizing Map (SOM) usually uses Euclidean distance to compare the samples of a class. But in this thesis, instead of using Euclidean distance, we use similarity measure to compare the feature vector of input digit and different neurons' weight. One or two samples are used to train the network and it doesn't need to have a set of samples in different fonts.

Keywords: OCR, SOM, similarity measure, feature extraction.



Shahrood University of Technology

Faculty Computer Engineering and Information Technology

The Recognition of Typed English Digits and Characters with Size and Font Limitation free through SOM Neural Network

Najmeh Samadiani

Supervisor:

Prof Hamid Hassanpour

July 2014